CONSTRAINT ON ABSOLUTE ACCURACY OF METACOMPREHENSION ASSESSMENTS: THE ANCHORING AND ADJUSTMENT MODEL VS. THE STANDARDS MODEL

By

HEEKYUNG KWON

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2011
To my parents
ACKNOWLEDGMENTS

I wish to thank several individuals for their help and support concerning my dissertation study. Primarily, I would love to express my gratitude towards Dr. Tracy Linderholm for being there through the most difficult times of my Ph.D. program. My research would not have been possible without the great support that she has been offering me for the last three years. I truly appreciate all the help that Dr. David Therriault offered me, particularly in the absence of my advisor. I am grateful that Dr. David Miller accepted my sudden request to be on my committee, and also for his advice on data analyses of the study. I also want to thank Dr. Stephen Pape for providing different perspectives on the study, which enriched my dissertation study.

I would like to thank other people, including my family, who have been a great help for me not only while conducting the study but also during the entire years of my Ph.D. program. First of all, I would like to thank Dr. Jongho Shin. The research training in his lab initiated my career as a researcher. He has continued to provide me with great advice and emotional support despite the long distance between us. I also thank Dr. David Painter, Shannon McCarthy, Jiyoung Kim, and Sangmin Shin for their kind help with the data collection of the study. Without them, it would have taken much longer to complete the study. I appreciate Heewon Kwon, my sister, for providing great support in many ways so that I was able to pursue this long journey towards a Ph.D. I am so lucky to have her as my sister. I am deeply grateful that I could have Hwimin Seo, my husband, right next to me during the last five years. I cannot even imagine my life in Gainesville without him. I also want to thank my parents-in-law for their prayers for me. Last, I am dedicating this dissertation study to my parents. This study is a small product of all the love that they have given me since I was born.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>7</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>8</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>9</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>11</td>
</tr>
<tr>
<td>Importance of Metacomprehension Assessments</td>
<td>11</td>
</tr>
<tr>
<td>Types of Measurements for Metacomprehension Accuracy</td>
<td>12</td>
</tr>
<tr>
<td>Three Findings Relevant to Absolute Accuracy</td>
<td>15</td>
</tr>
<tr>
<td>The Anchoring and Adjustment Model</td>
<td>20</td>
</tr>
<tr>
<td>Accounting for Limited Absolute Accuracy of Predictions</td>
<td>25</td>
</tr>
<tr>
<td>Accounting for Individual Differences in Absolute Accuracy of Predictions</td>
<td>27</td>
</tr>
<tr>
<td>Accounting for Superior Absolute Accuracy of Postdictions over Predictions</td>
<td>28</td>
</tr>
<tr>
<td>The Standards Model</td>
<td>30</td>
</tr>
<tr>
<td>Accounting for Limited Absolute Accuracy of Predictions</td>
<td>36</td>
</tr>
<tr>
<td>Accounting for Individual Differences in Absolute Accuracy of Predictions</td>
<td>41</td>
</tr>
<tr>
<td>Accounting for Superior Absolute Accuracy of Postdictions over Predictions</td>
<td>46</td>
</tr>
<tr>
<td>2 RESEARCH QUESTIONS AND HYPOTHESES</td>
<td>49</td>
</tr>
<tr>
<td>Question 1: Which Model Explains Limited Absolute Accuracy?</td>
<td>49</td>
</tr>
<tr>
<td>Question 2: Which Model Explains Individual Differences in Absolute Accuracy of Predictions</td>
<td>51</td>
</tr>
<tr>
<td>Question 3: Which Model Explains Superior Absolute Accuracy of Postdictions over Predictions</td>
<td>53</td>
</tr>
<tr>
<td>3 RESEARCH METHODS</td>
<td>55</td>
</tr>
<tr>
<td>Participants</td>
<td>55</td>
</tr>
<tr>
<td>Materials</td>
<td>56</td>
</tr>
<tr>
<td>Self-Perception of Reading Skill Task</td>
<td>56</td>
</tr>
<tr>
<td>Reading Skill Test</td>
<td>57</td>
</tr>
<tr>
<td>Comprehension Standards Task</td>
<td>59</td>
</tr>
<tr>
<td>Metacomprehension Accuracy Task</td>
<td>62</td>
</tr>
<tr>
<td>Modified Comprehension Standards Task</td>
<td>63</td>
</tr>
<tr>
<td>Procedure</td>
<td>64</td>
</tr>
<tr>
<td>Session 1</td>
<td>64</td>
</tr>
<tr>
<td>Session 2</td>
<td>66</td>
</tr>
</tbody>
</table>
4 RESULTS ..........................................................................................................................73

Which Model Explains Limited Absolute Accuracy of Predictions? ............................75
Which Model Explains Individual Differences in Absolute Accuracy of Predictions?......78
Which Model Explains Superior Absolute Accuracy of Postdictions over Predictions? .....79

5 DISCUSSION ..................................................................................................................86

Interpretations .....................................................................................................................86
  The Anchoring and Adjustment Model............................................................................86
  The Standards Model.........................................................................................................92
Implications and Limitations .............................................................................................98

APPENDIX

A A SAMPLE TEXT FOR THE READING SKILL TEST..................................................107
B SAMPLE QUESTIONS FOR THE READING SKILL TEST............................................109
LIST OF REFERENCES ......................................................................................................110
BIOGRAPHICAL SKETCH .................................................................................................118
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-1</td>
<td>Features of the texts used for the Reading Skill Test</td>
<td>69</td>
</tr>
<tr>
<td>3-2</td>
<td>Correlation coefficients of the comprehension standard (topic familiarity)</td>
<td>69</td>
</tr>
<tr>
<td>3-3</td>
<td>Features of texts used for the Metacomprehension Accuracy Task</td>
<td>69</td>
</tr>
<tr>
<td>3-4</td>
<td>Summary of the measures used for the current study</td>
<td>70</td>
</tr>
<tr>
<td>4-1</td>
<td>Descriptive statistics of individual variables</td>
<td>83</td>
</tr>
<tr>
<td>4-2</td>
<td>Descriptive statistics of the Metacomprehension Accuracy Task</td>
<td>83</td>
</tr>
<tr>
<td>4-3</td>
<td>Frequency data of absolute accuracy of predictions and postdictions</td>
<td>83</td>
</tr>
<tr>
<td>4-4</td>
<td>Absolute accuracy of predictions according to the level of the comprehension standard</td>
<td>83</td>
</tr>
<tr>
<td>4-5</td>
<td>Differences in absolute accuracy of predictions between more-skilled and less-skilled readers</td>
<td>84</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>An example of high relative- but low absolute accuracy (left) and low relative- but high absolute accuracy (right)</td>
<td>48</td>
</tr>
<tr>
<td>1-2</td>
<td>Hypothesized differences in the individual variance in the standard employed for predictions and postdictions</td>
<td>48</td>
</tr>
<tr>
<td>3-1</td>
<td>Six self-report questions for the Self-Perception of Reading Skill Task</td>
<td>71</td>
</tr>
<tr>
<td>3-2</td>
<td>Six self-report questions used to validate the Comprehension Standards Task</td>
<td>71</td>
</tr>
<tr>
<td>3-3</td>
<td>The procedure in which participants took the tasks and tests</td>
<td>72</td>
</tr>
<tr>
<td>4-1</td>
<td>Difference in self-perception of reading skill between more-skilled and less-skilled readers</td>
<td>85</td>
</tr>
</tbody>
</table>
The objective of this study is to provide a systematic account of three typical phenomena surrounding absolute accuracy of metacomprehension assessments: (1) the absolute accuracy of predictions is typically quite low; (2) there exist individual differences in absolute accuracy of predictions as a function of reading skill; and (3) postdictions are more accurate than predictions. To achieve the stated objective, I tested two models, the Anchoring and Adjustment model (Zhao & Linderholm, 2008) and the Standards model (based on Wiley et al., 2005) to determine which model better explains those three different findings. According to the Anchoring and Adjustment model (Zhao & Linderholm, 2008), when evaluating performance on a specific reading task, readers anchor around their self-perception of reading skill and make an adjustment away from it based on the perceived level of text understanding. In contrast, the Standards model, that I proposed by following Wiley et al. (2005)’s theory of metacomprehension assessment, states that readers set a standard based on their ideas of good comprehension and evaluate the perceived level of text understanding against the standard.

The study results suggested that the Anchoring and Adjustment model successfully explained all of the three findings in general. (1) The extent to which one’s self-perception of
reading skill deviated from reality accounted for readers’ limited absolute accuracy levels. (2) The different manner in which more- and less-skilled readers perceive their own reading skill accounted for individual differences in predictions between the two reading skill groups. (3) The magnitude of adjustments was greater for postdictions than predictions, as proposed by the Anchoring and Adjustment model, although self-perception of reading skill was more predictive of postdictions than predictions as opposed to the model’s hypothesis. Meanwhile, the data did not fit into the three accounts built from the Standards model’s standpoint overall. (1) The extent to which one’s standard of reading comprehension deviated from the test-giver’s standard did not account for limited absolute accuracy of predictions. (2) Differences in the standard of reading comprehension did not predict more-skilled readers’ underestimations and less-skilled readers’ overestimations of reading performance. (3) The new standards that readers built for postdictions were not closer to the test-giver’s standard than the original standards they formed for predictions, as opposed to the model’s prediction. Attempts were made to interpret findings contradicting each of the models’ hypotheses. Lastly, I discussed theoretical and educational implications, and limitations of the present study.
Metacognition consists of two subcomponents: one’s assessments of his or her own cognitive processes and one’s regulation of such operations (Flavell, 1979). Metacomprehension is a type of metacognition specifically relevant to the cognitive activity of reading comprehension. It refers to a reader’s capability of evaluating the quality of learning from text materials and regulating reading processes on the basis of the evaluation (Maki & Berry, 1984). For the last two decades, metacomprehension research has focused on the former in particular, that is, how readers accurately judge results of reading comprehension and predict performance on an upcoming test covering the text material (see Dunlosky & Lipko, 2007, for a review). There are two main reasons for the concentrated interest on the issue: First, accurate assessments provide the basis, at least in part, for effective regulation of reading processes, which results in productive learning from text materials (Son & Metcalfe, 2000; Thiede, Anderson, & Therriault, 2003). Accurate assessments of the extent of learning allow readers to efficiently distribute time and effort to different texts so that they can maximize the effectiveness of learning from them. Moreover, considering the fact that texts are a significant source of information and knowledge, and that reading comprehension is the core learning activity across all ages (e.g., Chapman & Tunmer, 1995; van den Broek, 2010), metacomprehension accuracy allows one to become an autonomous learner who takes responsibility for improving his or her knowledge. Second, despite such importance, metacomprehension accuracy has been found to be limited across a number of studies. Even college students, who are considered reasonably competent readers, normally over- or under-estimate their performance in comprehension tests (Glover, 1989; Grabe, Bordages, & Petros, 1990; Jacobson, 1990; Linderholm, Cong, & Zhao, 2008; Maki,
Shields, Wheeler, & Zacchilli, 2005). The averaged correlation coefficient indicating how well college students discriminate better-understood texts from less-understood ones was only around .27 across 25 studies conducted in Maki’s (1998a) laboratory. As one must understand why accuracy is low in order to propose how to improve it, many studies have investigated the mechanism underlying metacomprehension judgments and revealed the reasons for constraints on the accuracy (e.g., Dunlosky, Rawson & Middleton, 2005; Glenberg & Epstein, 1987; Griffin, Jee, & Wiley, 2009; Maki, Jonas, & Kallod, 1994; Morris, 1990). The current study is along the lines of those psychological efforts devoted to reveal the basis for constraints on metacomprehension accuracy and other relevant findings.

Types of Measurements for Metacomprehension Accuracy

Many studies have collected metacomprehension accuracy data using a similar method (e.g., Glenberg & Epstein, 1987). In the typical experimental setting, participants are asked to read several texts, and, after reading each text, predict how well they will perform in a subsequent comprehension test covering the text material they just read. In some studies, they are also asked to retrospectively evaluate their test performance after taking a comprehension test. Participants are normally instructed to predict or postdict their test performance in terms of percentage of correct answers (e.g., How well do you think you will perform in an upcoming comprehension test over the text just read?). Metacomprehension accuracy is computed on the basis of comparison between the estimated (predicted or postdicted) and the actual test performance over all the texts given in each study.

Making a prediction seems to take a different mental route from making a postdiction (e.g., Griffin, Jee, & Wiley, 2009; Zhao & Linderholm, 2008). One critical difference between the two occasions is whether people have taken the criterion test or not. Accordingly, certain types of cues used to make metacomprehension assessments (e.g., feelings about the difficulty of the
criterion test) are available only for postdictions. Thus, predictions and postdictions need to be investigated in a separate manner.

For the last few decades, the accuracy of predictions has received far more academic attention than postdictions (see Kwon & Linderholm, in press, for a review). The focused interest may be due to the greater practical implication of prediction accuracy relative to postdiction accuracy. Accurate predictions can be used as the basis for judgment regarding how much further one needs to study a text, which helps readers to improve reading comprehension before they have to be tested. In contrast, accurate postdictions only provide post-hoc information on how well they performed on the comprehension tests. Thus, it is not necessary that readers with high and low postdiction accuracy demonstrate differences in effectiveness in distributing their time and effort to different texts before taking comprehension tests covering them. Hence, accuracy of postdictions is not as useful as predictions in terms of enhancing reading achievement. Additionally, studies have reported that predictions are less accurate than postdictions even if the latter are still far from perfect (Hacker et al., 2000, Maki et al., 2005; Miesner & Maki, 2007; see also Connor, Dunlosky, & Hertzog, 1997, for other domains). For these two reasons, many studies, including the current one, have their main focus on constraints of prediction accuracy rather than those of postdiction accuracy.

There are two types of accuracy for both prediction and postdiction, which is determined by computational methods: absolute and relative accuracy (e.g., Glenberg & Epstein, 1987; Hacker et al., 2000). Absolute accuracy refers to the extent to which specific assessments match with actual test performance (Maki, 1998b). There are different measures for absolute accuracy (Maki, 1998b; Mengelkamp & Bannert, 2010), and the most commonly used indicator is the so-called bias measure. This is the signed difference between mean prediction (or postdiction) and
mean actual test performance (e.g., Maki, 1998b; Hacker et al., 2000; Maki et al., 2005). The sign indicates whether the reader over- or under-estimated his performance. If an individual’s assessments are above his actual test performance on average (e.g., mean assessment is 82% and mean performance is 60%), the person is said to overestimate his performance and the absolute accuracy value has positive sign (e.g., +22%). In contrast, an individual is said to underestimate his performance if he performs better than he has predicted or postdicted. In that case, absolute accuracy has a negative value. Importantly, the absolute value, regardless of its sign, indexes the extent to which one’s absolute accuracy is limited. That is, the larger the gap between assessed and actual performance the lower absolute accuracy is.

Meanwhile, relative accuracy refers to the extent to which assessments are correlated with performance across texts (Nelson, 1984). To illustrate, if an individual tends to make high assessments over texts in which test performance is high and makes low judgments over texts with low test performance, he is said to have a high relative accuracy. Inversely, if a reader makes high assessments for texts in which test performance are comparably poor and makes low assessments on texts whose performance is high, he is said to have a low relative accuracy. In other words, if a reader’s assessments fluctuate in accordance with change in the actual performance, he or she is seen to display a high relative accuracy. Therefore, the mean correlation coefficient indicates how accurately a reader discriminates well-comprehended texts from poorly comprehended ones. The values range from -1, indicating that a reader’s judgments of his performance change in the exact opposite way of the actual test performance results, 0 indicating that there is absolutely no discrimination in assessments regardless of actual performance, to +1 implying that assessments are in accord with the actual performance across the given texts.
Three Findings Relevant to Absolute Accuracy

As mentioned previously, both absolute and relative accuracy of predictions are low even among college students. Regarding absolute accuracy of predictions, they likely over- or underestimate their performance in comprehension tests to a varying degree (Glover, 1989; Grabe, Bordages, & Petros, 1990; Jacobson, 1990; Linderholm, Cong, & Zhao, 2008; Maki et al., 2005). Likewise, relative accuracy of predictions is only around .27 on average across many studies (see Maki, 1998a, for a review). Noticeably, a great deal of academic attempts have been made to explain unimpressive relative accuracy (e.g., Baker & Dunlosky, 2006; Dunlosky, Rawson, & Mcdonald, 2002; Glenberg et al., 1987; Maki & Serra, 1992; Morris, 1990, see also Dunlosky & Lipko, 2007, for a review). Many researchers have agreed upon the perspective of the cue-utilization theory for explaining low relative accuracy (e.g., Dunlosky & Rawson, 2005; Griffin, Jee, & Wiley, 2009; Koriat, 1997). According to this theory, low relative accuracy is attributable to the use of inappropriate cues such as familiarity of text topic (e.g., Glenberg & Epstein, 1987; Glenberg et al., 1987), momentary retrieval of merely a few concepts (e.g., Dunlosky, Rawson, & Mcdonald, 2002; Dunlosky, Rawson & Middleton, 2005; Morris, 1990), or a rough estimation of ease of text processing (e.g., Rawson & Dunlosky, 2002) as the basis for metacomprehension predictions. To explain, readers seem to infer the extent to which they have comprehended a text from a certain cue or several cues that are accessible to them at the moment of being asked to make a prediction. They then reach a conclusion about how well they will perform on an upcoming comprehension test covering the text (e.g., Koriat, 1997). However, such cues are unlikely to be valid or reliable as an indicator of the extent of text understanding (e.g., Topic familiarity does not reliably correlate with the extent to which the text is understood). This is considered to be the main reason for the limited relative accuracy of predictions among researchers.
As opposed to relative accuracy, studies have rarely been conducted to account for low absolute accuracy, even though a few studies have investigated relationships between some factors (e.g., reading skill, text difficulty) and absolute accuracy (e.g., Maki et al., 2005). One can argue that cue invalidity can cover limited absolute accuracy as well as relative accuracy. However, in the current study, I suspect that an exclusive reason exists for absolute accuracy separate from cue invalidity from both a theoretical and an empirical stance. First of all, it is theoretically plausible that one displays a high relative accuracy and a low absolute accuracy simultaneously. To elaborate, if one’s predictions fluctuate in accordance with actual performance, he would display a high relative accuracy even when those predictions are made far from the actual performance, leading to a low absolute accuracy (Figure 1-1). The opposite case is also possible. That is, one would have a high absolute accuracy as long as the mean prediction approximates the mean actual performance even if his judgments made are insensitive to the fluctuation in the actual performance, which yields a low relative accuracy (Figure 1-1).

Empirical data support this divergence between the two types of accuracy as well. One study reported a weak correlation between absolute and relative accuracy ($r = -.027$) but also reported a strong correlation between two absolute accuracy scores of predictions and postdictions ($r = .832$), and between those two relative accuracy scores ($r = .316$) (Maki et al., 2005). Additionally, college students tend to be consistent in terms of the level of absolute accuracy across occasions (Schraw & Roedel, 1994; West & Stanovich, 1997), whereas their relative accuracy fluctuates over time (Kelemen, Frost, & Weaver, 2000). That is, one who displays a certain level of relative accuracy on one occasion shows a fairly different level of relative accuracy one another. These findings strengthen the theoretical supposition that there is a reason for limited absolute accuracy separate from relative accuracy, even if it is uncertain
whether cue invalidity affects the former as well as the latter. Therefore, I argue that, asides from the cue invalidity, investigations are needed to reveal the constraint exclusively operating on absolute accuracy.

Besides the lack of academic efforts to reveal the reason for limited absolute accuracy, two empirical findings pertaining to absolute accuracy also have yet to be explained. First, previous studies have consistently reported that more-skilled readers are likely to be conservative while less-skilled readers tend to be overly generous in evaluating their reading performance (e.g., Glover, 1989; Grabe, Bordages, & Petros, 1990; Jacobson, 1990; Linderholm et al., 2008; Maki et al., 2005). To reiterate, it seems that, when making predictions, readers who achieve high scores in various reading comprehension tests tend to underestimate or are less likely to overestimate than readers who achieve low scores in those kinds of tests. Until now, notably, no account has been provided for this difference in absolute accuracy of predictions between more- and less-skilled readers. Another interesting finding regarding absolute accuracy is that postdictions are likely to be more accurate than predictions (Hacker et al., 2000, Maki et al., 2005; Miesner & Maki, 2007; see also Connor, Dunlosky, & Hertzog, 1997, for other domains). This finding also remains to be explained psychologically. It should be noted that academic efforts have been made to understand increases in relative accuracy from predictions to postdictions (e.g., Maki, 1998b; Pierce & Smith, 2001) but not for absolute accuracy. Overall, based on the notion of the psychological independence of absolute accuracy from relative accuracy and the relative academic indifference about it, I have attempted to account for the following three phenomena in this study: (1) the limited absolute accuracy of predictions, (2) individual differences in absolute accuracy of predictions as a function of reading skill, and (3) the superior absolute accuracy of postdictions over predictions.
A close investigation of the psychological route to metacomprehension assessments provides a systematic explanation of these three phenomena. I suppose that low absolute accuracy of predictions originates from a limitation residing in the procedure taken to reach them. Likewise, a certain difference in the procedure readers use to arrive at a metacomprehension prediction is also thought to cause the variation between more- and less-skilled readers. Additionally, as briefly discussed earlier, postdictions are made through a different mental route from predictions, which is suspected to be the cause of the superior accuracy of the former. Unfortunately, however, not enough attention has been paid to the entire procedure that readers go through when making metacomprehension assessments. It appears that many researchers agree that, during the procedure, the extent to which a specific text is understood is perceived based on more or less invalid cues (e.g., Baker & Dunlosky, 2006; Maki & Serra, 1992). However, from some theoretical suggestions on the mental route, which will be introduced in the next section in detail, it is inferred that perception of the extent of understanding is just a part of the procedure. Additionally, inaccuracy of the perception due to cue invalidity is only enough to explain limited relative accuracy. In order to understand low absolute accuracy and the relevant phenomena, it seems necessary to investigate what other steps are taken in the procedure readers go through to make metacomprehension assessments in addition to the perception of the extent of text understanding. I suspect that an examination of the unknown steps will provide tips to understand the phenomena surrounding absolute accuracy.

There are two different models proposed for the mental procedures underlying metacomprehension predictions. First, according to the Anchoring and Adjustment model (Zhao & Linderholm, 2008), when making a prediction, readers anchor around their self-perception of reading skill and make an adjustment away from it based on the perceived extent to which they
comprehended a certain text. Meanwhile, in the present study, I propose the Standards model based on theoretical suggestions Wiley and colleagues (2005) made regarding the mental procedures of metacomprehension assessments. According to Wiley et al., when assessing their reading performance, readers set a ‘standard’ based on their concept of good comprehension and evaluate the perceived extent of text understanding against it. I will use the basic tenets of these models to address the aforementioned unanswered questions about metacomprehension assessments.

To compare these two models briefly, they commonly include the perception of the extent to which a specific text is comprehended as a step of the procedure of metacomprehension prediction. In the Anchoring and Adjustment model, it provides the basis for the magnitude of adjustment made away from the anchoring point. In the Standards model, it is viewed to be evaluated against the standard that readers form. However, each model proposes a different additional step: ‘anchoring’ for the Anchoring and Adjustment model and ‘standard-setting’ for the Standards model. Noticeably, a different individual variable plays a critical role in the process of making metacomprehension assessments: one’s self-perception of reading skill versus one’s concept of good comprehension. Both of the individual variables are viewed to be crucial personal traits that affect one’s reading performance (Chapman & Tunmer, 1995; 1997; Chapman, Tunmer, & Prochnow, 2000; Stringer & Heath, 2008; Valentine, Dubois, & Cooper, 2004, for self-perception of reading skill, Schommer, 1990; Schommer, et al., 1992, for theories of personal beliefs about good comprehension). For instance, children who define themselves a good reader tend to be motivated to read more frequently and such active involvement contributes to improving their reading ability (e.g., Chapman, Tunmer, & Prochnow, 2000). Thus, it is inferred that these individual variables have to do with readers’ assessments of their
reading performance. However, no academic approach has been made to supply a comparable account for how each of the variables is involved in metacomprehension assessments. Thus, this study will provide initial empirical evidence in terms of which of them has a greater impact on the metacomprehension assessment by revealing which of the two models is true about the entire procedure. For the purpose, I will investigate which model comprehensively covers the unanswered three phenomena relevant to absolute accuracy: (1) limited absolute accuracy of predictions (2) individual differences in absolute accuracy of predictions as a function of reading skill, and (3) superior absolute accuracy of postdictions over predictions. Primarily, I will get into an in-depth discussion regarding how each of the Anchoring and Adjustment model and the Standards model accounts for them throughout the next two sections.

**The Anchoring and Adjustment Model**

Assuming that readers attempt to predict their performance as accurately as possible, what is the optimal psychological procedure that they should follow to achieve the goal? It would seem desirable for them to choose a text of a certain level of difficulty as a criterion. They could then compare the current text to the criterion text in terms of the difficulty. In order to make an accurate prediction of performance on a subsequent comprehension test covering the current text, they should use their test performance on the criterion text as the anchoring point and make adjustments away from it. These adjustments can be based on accurate perception of the extent to which they comprehended the text relative to the understanding level of the criterion text.

However, it does not seem that readers follow such an ideal procedure to arrive at accurate predictions for several reasons. For instance, it appears very difficult for readers to choose a criterion text out of numerous texts that they have read, and to recall the exact level of performance that they attained in the comprehension test covering that text material. Accuracy of metacomprehension predictions has been fairly low throughout several studies (e.g., Glover,
1989; Maki, 1998a), which supports the notion that readers do not employ such an ideal procedure when making metacomprehension predictions.

What is the reality of the procedure that readers use to make estimates of comprehension? According to the Anchoring and Adjustment model (Zhao & Linderholm, 2008), one takes two steps: (1) anchoring on self-perception of reading skill and (2) making an adjustment away from it based on their perception of the extent to which they comprehended the text that they just read (see also Scheck, Meeter, & Nelson, 2004, for other domains). Note that this process proposed is similar to the optimal one discussed above: both posit that readers anchor on a certain point and make an adjustment away from it. However, there exist two big differences between the optimal procedure and the presumably virtual procedure that the Anchoring and Adjustment model proposes. First of all, the Anchoring and Adjustment model suggests that the actual anchoring point employed is a subjective, self-perception of reading skill, not an objective level of the performance that the reader displayed in a comprehension test (Zhao & Linderholm, 2008). According to this model, readers have formed an idea of how good of a reader they are based on their history of reading texts and being exposed to comprehension tests. Their assessments start with conscious or unconscious retrieval of such a subjective idea of their own reading skill rather than relying on objective level of their reading performance.

Next, the Anchoring and Adjustment model claims that adjustments are unlikely to be sufficient. This opposes the optimal model, which supposes that an adequate amount of adjustment is made based on the perceived difference of understanding between the criterion and the current text. Zhao and Linderholm (2008) have mainly ascribed insufficiency in the adjustment to uncertainty about the subsequent comprehension test (See also Maki, 1998a). Since readers have to make predictions without the benefit of knowing what kinds of questions
will be on subsequent tests, they would rather choose a secure, conservative way of sticking to
the anchoring point than risking an adjustment of greater magnitude in accordance with the
fluctuating extent of text understanding. In addition to the test uncertainty, it is inferred that the
cue invalidity contributes in yielding insufficient adjustments as well. Even if Zhao and
Linderholm have not specified the impact of invalid cues on the magnitude of adjustments, they
have claimed that adjustments are made to the perceived extent of text understanding. According
to the cue utilization theory that this model agrees with (Zhao & Linderholm, 2008), such
perceptions take place based on somewhat invalid cues (e.g., familiarity of text topic) available
at the moment of making predictions. Thus, cue invalidity as well as test uncertainty should
influence insufficiency in the adjustments. I claim that readers plausibly appreciate invalidity of
the cues to a varying degree, and such appreciation works to reduce their confidence in making a
sufficient amount of adjustments on that basis. However, I consider that these two factors differ
from each other in terms of the manner that each affects adjustments. It is suspected that test
uncertainty mainly functions to diminish the magnitude of adjustments by lowering readers’
confidence about their perception of the extent of text understanding. Comparably, cue invalidity
influences both the magnitude and the direction of adjustments because it not only hurts this
confidence but also confuses them regarding which text is more or less understood than another.
In sum, the current study claims that test uncertainty and cue invalidity combine to exert an
impact on the adjustments insufficiently made away from the anchoring point set around self-
perception of reading skill.

Putting these two features together, if the Anchoring and Adjustment model is a realistic
account of metacomprehension assessment procedures, metacomprehension predictions would
end up being near the anchoring point due to insufficient adjustments made away from the
anchoring point. In other words, predictions are mainly affected by their self-perception of reading skill rather than fluctuation in the extent of comprehension across the texts given. Thus, if one considers himself a good reader, he tends to predict that his performance will be high in general. Inversely, one who regards himself as a poor reader, his predictions will likely be low, regardless of fluctuation in the text difficulty. Empirical data support this large dependence of metacomprehension predictions on self-perception of reading skill. Studies across domains outside reading comprehension have reported that performance assessments strongly correlate with self-perceptions of ability (see Ehrlinger & Dunning, 2003, for a review). Studies on metacomprehension assessments provided evidence to support the insufficient adjustments: they reported no significant variation among predictions that an individual reader made even when the difficulty varied across the texts given (Hacker et al., 2000; Moore, Lin-agler, & Zabrucky, 2005). Additionally, in Zhao’s (2008) study, students rated their self-perception of reading skill on a 5-point Likert scale with 1 indicating ‘very poor’ and 5 indicating ‘excellent.’ The self-perception of reading skill was found to be predictive of predictions even after controlling for the actual performance. Also, she showed that predictions fluctuated with text difficulty but only to a lesser degree than fluctuation of the actual performance. All of these study results support that metacomprehension predictions are largely dependent on self-perception of reading skill as the Anchoring and Adjustment model hypothesizes.

In order to test this model, therefore, one needs to identify what constitutes one’s self-perception of reading skill. Unfortunately, no studies have investigated adult readers’ self-perception of reading skill, although many have been carried out on children’s self-perception of reading skill (e.g., Bouffard et al., 2003; Chapman & Tunmer, 1997; Harter, 1982; Helmke & van Aken, 1995; Henk & Melnick, 1995). Through the current study, hence, I propose the
construct of adult readers’ self-perception of reading skill, which is presumed to have a critical impact on their metacomprehension predictions. Primarily, I claim that self-perception of reading skill consists of different perceptions over varying sub-skills involved in on-line processes of reading. On the basis of reading experience accumulated, it is highly possible that adult readers conceive the fact that text processing is a complicated cognitive process involving a variety of sub-skills from lower order, basic processes (e.g., phonological encoding, syntactic parsing) to higher order processes (e.g., knowledge-based interpretation, making connections between different pieces of text information) (e.g., Kintsch, 1998). Particularly, college students probably perceive that higher order processes play a more critical role in rendering individual differences in reading performance, as many of them are capable of fluently dealing with lower order processes such as decoding, while occasionally struggling with higher order ones such as inference-making (Perfetti, 1985). Most of all, a large proportion of lower order processes tend to take place in an automatic manner while higher order ones are more likely to occur strategically (e.g., Walczyk, 2000). Thus, the latter is more prone to be consciously captured by readers than the former. In regard to the off-line product of reading, additionally, adult readers possibly view the ability of remembering text ideas as a critical skill required for successful reading comprehension because numerous reading comprehension tests that they have taken require them to retrieve what they learn from text materials. I also infer that they have formed a perception of their reading speed as an indicator of reading performance based on experiences that easy texts read more fluently than difficult ones (Lorch, Lorch, & Matthews, 1985; Zwaan, 1996). Taken together, the current study suggests that adult readers’ self-perception of reading skill consists of separate perceptions of higher order sub-skills of reading comprehension,
memory ability and reading speed. It is expected that such perceptions jointly exert a large influence on metacomprehension predictions if the Anchoring and Adjustment model is the case.

**Accounting for Limited Absolute Accuracy of Predictions**

Zhao and Linderholm (2008) have identified two distinct sources for limited accuracy of metacomprehension predictions. One is uncertainty about the criterion test subsequently given, and the other is the poor diagnostic validity of the cues used. As explained previously, both exert a decreasing influence on the accuracy of the adjustments made away from the anchoring point in terms of the magnitude and the direction. They have argued that inadequate adjustments mainly constrain *relative* accuracy. I agree with them because they prevent predictions from accurately reflecting fluctuation in the extent of understanding among texts. Zhao and Linderholm have viewed the inadequate adjustments as a constraint for *absolute* accuracy as well. It seems to make sense because they may affect how far predictions are away from the actual performance. Setting aside examining impact of the inadequacy in adjustments on absolute accuracy, however, I speculate that there should be an exclusive reason for absolute accuracy. As explained earlier, empirical data (e.g., a low correlation between the two types of accuracy) implies separation of the reason for limited absolute accuracy from relative accuracy. Theoretically, absolute accuracy can be high as long as the averaged prediction approximates the mean actual performance even when their fluctuations do not correspond. To rephrase this tenet from the Anchoring and Adjustment model’s perspective: even when the adjustments are improperly made away from the anchoring point, the mean prediction can be close to the mean actual performance but only if the anchoring point is set appropriately. Since this model claims that predictions are biased toward the anchoring point due to insufficient adjustments, it is crucial where readers set the anchoring point in order to make accurate predictions on average. However, the position of the anchoring point is viewed to have no big impact on the sensitivity
to fluctuation of the extent of text understanding because the anchoring point is supposedly fixed for every adjustment made on one occasion. Overall, whereas inadequate adjustments might work to lower both absolute and relative accuracy, inappropriate placement of the anchoring point seems to be an exclusive reason for limited absolute accuracy.

In order to specify the constraint on absolute accuracy separate from relative accuracy, thus, one needs to closely examine where readers set their anchoring point, and explain how the location can be inappropriate. As introduced previously, the Anchoring and Adjustment model claims that the anchoring point is normally set around self-perception of reading skill (Zhao & Linderholm, 2008). Because of the insufficient adjustments, metacomprehension predictions are anticipated to largely depend on one’s self-perception of reading skill according to this model. In regard to the appropriateness of the anchoring point, note that studies outside the area of reading comprehension have reported that people tend to have a distorted self-perception of competence across domains (see Dunning, Heath, & Suls, 2004, for a review). Zhao and Linderholm have inferred, from the great reliance of predictions on self-perception of reading skill, that inaccurate self-perceptions have a negative impact on absolute accuracy. Unfortunately, however, not enough empirical data have been attained on how accurately people perceive their own reading skill. Additionally, no attempt has been made to reveal the impact of inaccurate reading self-perception on absolute accuracy, let alone to provide an account for it.

In the present study, I try to explain how inaccurate self-perception of reading skill can decrease absolute accuracy. First of all, I note that one’s objective level of reading skill is known as a reliable predictor of his or her performance in reading comprehension tests on specific occasions (e.g., Perfetti, 1985). Virtually, one’s actual reading skill is determined based on results from a number of reading comprehension tests one has taken. Additionally, self-
perception of reading skill is expected to be predictive of metacomprehension predictions as discussed above. Remember that absolute accuracy is defined as discrepancy between the mean predicted and the mean actual performance in reading comprehension tests. Thus, the more a reader’s self-perception of reading skill deviates from the objective level of reading skill, the greater the gap is expected between actual and predicted performance, that is, the lower the absolute accuracy. Consequently, the inappropriately set anchoring point based on inaccurate self-perception of reading skill is viewed to be the critical, exclusive reason for limited absolute accuracy. Therefore, I hypothesize that the gap between subjectively perceived and objectively assessed reading skill level accounts for absolute accuracy.

**Accounting for Individual Differences in Absolute Accuracy of Predictions**

It is time to discuss how the Anchoring and Adjustment model explains individual differences in absolute accuracy of predictions found in previous studies (Glover, 1989; Grabe, Bordages, & Petros, 1990; Jacobson, 1990; Linderholm, Cong, & Zhao, 2008; Maki et al., 2005). In those studies, more-skilled readers were less likely to overestimate their reading performance than their less-skilled counterparts. In other words, more-skilled readers’ predictions are generally lower or slightly higher than the actual performance in comprehension tests. However, the less-skilled readers’ predictions are normally higher than the actual performance to a great extent. Assuming the erroneous self-perception of reading skill as a critical reason for limited absolute accuracy, I suspect that the difference between the two reading-skill groups might derive from the different manner in which each group perceives their reading skill. My hypothesis is that more-skilled readers are likely strict in terms of perceiving their own reading skill while less-skilled readers tend to be overly generous about their reading ability. Previous studies in domains outside reading comprehension support the hypothesis by demonstrating that less-skilled people tend to exaggerate their own ability to a larger extent than more-skilled ones.
(e.g., Ditto & Lopez, 1992; Dunning, 2004; Kunda, 1990; Miller & Ross, 1975). If this is the case with reading comprehension as well, it is expected that more-skilled readers anchor at a point that is lower or slightly higher than their actual level of reading skill, and then make just a small amount of adjustments. Therefore, their performance would be likely underestimated or overestimated to a small degree. In contrast, less-skilled readers are thought to hold a comparably too optimistic perception of their reading skill. Thus, their anchoring points would likely be far higher than the objective level of reading skill, which results in a large magnitude of overestimations; it is also assumed that they tend to make trivial adjustments away from it.

In order to provide evidence that the Anchoring and Adjustment model covers this individual difference as a function of reading skill, I developed a measure of each group’s self-perception of reading skill based on the construct of adult readers’ self-perception, as explained previously. I then compared it with their objectively measured reading performance. If more-skilled readers’ self-perception of reading skill tends to be lower than the objective reality, while less-skilled readers’ self-perceptions of reading skill are greatly higher than it actually is, it can be concluded that the Anchoring and Adjustment model captures individual differences in absolute accuracy as a function of reading skill. Again, this is assuming that inaccurate self-perception of reading skill is the critical reason for low absolute accuracy.

**Accounting for Superior Absolute Accuracy of Postdictions over Predictions**

Previous studies have shown that postdictions retrospectively made on performance on past comprehension tests are more likely to be accurate than predictions prospectively made before taking such tests (Hacker et al., 2000; Maki et al., 2005; Miesner & Maki, 2007). This divergence in accuracy implies that there exist differences in the mental procedures taken between these two types of assessment (Griffin, Jee, & Wiley, 2009; Pierce & Smith, 2001). Remember that a large dependency on inaccurate self-perception of reading skill is suspected to
be the exclusive cause for limited absolute accuracy of predictions. Such a great dependency derives from readers’ tendency of making only a trivial magnitude of adjustment away from the anchoring point. Therefore, it is inferred that improved accuracy is attributable to decreased dependency on self-perception of reading skill due to increased adjustments when postdictions are made compared to predictions (Zhao & Linderholm, 2008).

To explain specifically, one of the most critical reasons for such insufficient adjustments is that readers are uncertain about the upcoming comprehension test (e.g., types or difficulty of the testing items) when being asked to predict their performance. However, Zhao and Linderholm (2008) have noted that test uncertainty largely decreases on the occasions of making postdictions since they have already taken the criterion test. Meanwhile, cue validity is enhanced when making postdictions (Griffin, Jee, & Wiley, 2009). Some cues that are unavailable for predictions are now accessible when postdictions are made (e.g., feeling about the test difficulty). These new cues generally have a superior diagnostic validity of reading performance compared to the cues that exist for predictions (e.g., topic familiarity). I argue that readers not only conceive limited diagnostic validity of the cues for predictions, but also appreciate such improvement in the cue validity for postdictions. Taken together, not only decreased test uncertainty but also an increase in the cue validity allows one to be more confident about making adjustments in accordance with fluctuations in the text difficulty after having taken the criterion test. Therefore, it is expected that the variance in the adjustments across the texts given is greater for postdictions than predictions (Zhao & Linderholm, 2008). Accordingly, dependency on inaccurate self-perception of reading skill is predicted to be smaller for postdictions than predictions. Finally, the gap between perceived and actual reading skill is hypothesized to have a smaller explanatory power for accuracy of postdictions than that of predictions. Zhao (2008)
provided evidence for an increase in the adjustments for postdictions relative to predictions. She showed that postdictions changed as a function of text difficulty to a greater extent than predictions. Consistently, in her study, self-perception of reading skill correlated with predictions more strongly than postdictions, controlling for actual performance. This would indicate a lower dependency of postdictions on the anchoring point than predictions.

Until present, I have explained how the Anchoring and Adjustment model accounts for the three different findings relevant to absolute accuracy. Throughout the following section, I will precisely examine how the Standards model views each of them in the same manner.

**The Standards Model**

I propose an alternative model, which is called the Standards model, as the psychological route that readers go through when making metacomprehension predictions. Like the Anchoring and Adjustment model, this model proposes that one takes two steps: (1) perceiving the extent to which the current text is comprehended and (2) evaluating it against a standard. In this model, following Wiley et al.’s (2005) suggestion, readers are viewed to employ a standard instead of an anchoring point, against which specific reading performance is evaluated. According to Wiley et al. (2005), readers have a different idea of good comprehension, which is consistent with the Landscape model of text processing (van den Broek et al., 1996). The Landscape model proposes that the so-called Standard of Coherence serves as a guide for on-line processes of reading. In comparison, Wiley et al. have suggested that such a concept works as the standard one will set for metacomprehension predictions. They made no explicit comments regarding how one perceives the extent to which specific texts are comprehended before it is evaluated in reference to the standard. However, they seemed to agree with the cue utilization theory as the Anchoring and Adjustment model does. Thus, the Standards model claims that the extent of text understanding is perceived based on available cues. Overall, this model proposes that the
procedure underlying metacomprehension prediction includes setting a standard based on one’s concept of good comprehension and evaluating the perceived extent of text understanding against it.

There may be several ways in which one’s idea of good comprehension is conceptualized. Wiley et al. (2005) have suggested one way based on Kintsch’s (1998) theory of the levels of text representation (see also van Dijk & Kintsch, 1983). According to the theory, a text can be represented in the following three levels as a result of reading comprehension: the surface model, the text-base, and the situation model. The surface model is the most superficial level of representation. Encoding of lexical and syntactic dimensions of a text produces this level of text representation. Readers, adult readers in particular, rarely pursue building a surface model as their reading goal. Semantic analysis of a text yields the text-base. In a text-base, thus, text ideas are rearranged in propositional forms but the global structure of the text is preserved. In order to build a good text-base, one needs to reproduce as many text ideas as the author has presented on the text surface. The situation model is built when readers try to represent not only the text content itself but also the whole situation that it is placed in. To construct this type of model, therefore, readers need to entirely reorganize text ideas in the way that makes sense to them and interpret those ideas based on their background knowledge (Zwaan & Radvansky, 1998, see also Fletcher, Chryser, 1990; Radvansky, Zwaan, & Curiel, 2001; Schmalhofer & Glavanov, 1986, for empirical evidence). Overall, one’s idea of good comprehension can be placed somewhere on the continuum between one end of the construction of a text-base and the other end of the building of a situation model.

I claim that the situation model level of text representation is higher standard of good reading comprehension, in general, than the formation of a text-base level of text representation.
Reading comprehension involves both memorization of text ideas, which is critical to build a good text base, and in-depth understanding of them, which is central for construction of a good situation model (e.g., Kintsch, 1994). Strictly speaking, however, comprehension means obtaining the latter going beyond the former, among reading researchers (e.g., Kintsch, 1994; Rawson, Dunlosky, & Thiede, 2000; Wiley et al., 2005). It is because, whereas memorization of text ideas only allows readers to reproduce the text (e.g., recall), in-depth understanding of the subject matter enables the text information to be applied even in completely novel situations (e.g., problem-solving) (e.g., McNamara et al., 1996; Moravcsik & Kintsch, 1993). Moreover, a good situation model is built on the basis of a good text-base constructed at least to a certain extent (Kintsch, 1994; McNamara et al., 1996). For instance, text-ideas can be well-elaborated with background knowledge only after their meanings are successfully represented. Millis and colleagues (1998) have supported the argument by showing that, during the first trial of reading a difficult text in particular, mental resources mainly go to building a text-base while a small amount of them, if any remains, are devoted to form a situation model. By contrast, when reading the same text for the second trial, readers are able to pay greater attention to situation model-building due to a reduced burden of representing text ideas themselves. The study results show that the text-base construction is a sort of prerequisite for the formation of a situation model. Additionally, it appears that the former is generally easier to build than the latter. The text-base construction requires a certain amount of local-level connecting activities to transform words to meaningful, propositional units (McNamara et al., 1996). By contrast, in order to build a situation model, one needs to trigger considerable knowledge-based inferences not only to fill in what is unstated in the text, but also to reorganize the text ideas at a global level (McNamara et al., 1996). Thus, building a situation model seems more resource-consuming than building a
Overall, the situation model building is a more ultimate level of reading purpose, which can be pursued after attaining the goal of text-base construction. Consistently, in the Standards model, it is assumed that readers with a low standard are likely satisfied with building a text-base representation and limited in noticing the value of building a situation-model as a reading purpose. In contrast, readers with a high standard are said to aim at building a situation model beyond formation of a text-base.

On the basis of the conceptualization of the levels of the standard used for metacomprehension predictions, this study proposes a method to measure one’s standard about what entails good reading comprehension. Suppose that readers are presented with a set of memory-based and inference-based questions, both of which are typically used to evaluate one’s reading performance for academic purposes (e.g., Rawson, Dunlosky, & Thiede, 2000). Memory-based questions refer to testing items used to evaluate readers’ memory of somewhat discrete facts or details explicitly presented on the text surface. In contrast, inference-based questions are used to evaluate whether readers obtain in-depth understanding of the text by reorganizing texts ideas or integrating them with background knowledge. Consider that they are asked to rate each of those questions in terms of how appropriate it is as a comprehension testing item. It is predicted that readers with a low standard will rate memory-based questions higher than inference-based questions. It is because, due to the limitation in valuing in-depth understanding of the principles or concepts underneath the text content, their concept of good comprehension is to build a text base representing text ideas as the author intended on the text surface. Oppositely, readers with a high standard are expected to value inference-based questions more than memory-based questions in accordance with their idea of good comprehension. They believe that readers need to make a variety of knowledge-based inferences rather than memorize
what the text displays on the surface to attain their goal of reading, that is, construction of a situation model. In conclusion, I propose that one’s reading standard can be indexed by the extent to which he values memory-based questions versus inference-based questions as a good comprehension-testing item.

Previous studies have shown that readers use a different type of standard in terms of evaluating their reading performance (e.g., Baker, 1985; Muis, 2007; Ryan, 1984). Among those studies, I pay attention to Ryan (1984) reporting that readers’ *epistemological beliefs* affect their standard applied to assessments of their reading performance. Notably, the standard reflecting personal epistemology appears to have the same concept as what the Standards model proposes. To explain, personal epistemology refers to one’s ideas or beliefs about the nature of knowledge and knowledge acquisition (e.g., Schraw, 2000). One’s epistemological beliefs can be placed somewhere on the continuum with one end indicating naïve theorists and the other end indicating sophisticated theorists. Naïve theorists believe that knowledge is a simple accumulation of isolated facts, and is directly conveyed from experts to novices with no change in the meaning. In contrast, sophisticated theorists hold beliefs that knowledge consists of interconnected concepts; and learning is a procedure of meaning construction because it is transformed throughout personal interpretation when it is conveyed from one person to another. Note that, in Ryan (1984)’s study, when subjects were asked what standard they used to determine how well they comprehended the text materials given, naive theorists reported using memory-oriented standards (e.g., how much text information they can recall) while sophisticated theorists showed preference for inference-oriented standards (e.g., to what extent they can apply text information to new situations). The interpretation of the study results is that naïve theorists consider memorization of what is stated in the text as the most important activity involved in learning
from texts. On the contrary, sophisticated theorists are viewed to value different types of
inference-making (e.g., integrations of text ideas, interpretation of text information based on
background knowledge) more than memory-based processes when it comes to achieving their
idea of desirable comprehension. According to the operational definition explained previously,
readers with a low standard refer to those who value memory-based questions over inference-
based questions as a good comprehension testing item, and the opposite is the case for readers
with a high standard. Therefore, those with lower standards can be considered naïve theorists
who view learning from texts as a transmitting process of knowledge from authors to readers. In
contrast, those with higher standards correspond to sophisticated theorists who value readers’
personal meaning construction during reading as a process of knowledge acquisition.
Consequently, the standard of the desirable comprehension conceptualized in this model is said
to reflect individual readers’ epistemological beliefs.

Assuming the correspondence between the level of personal epistemology and that of the
standard of good comprehension, study results on personal epistemology can be used as evidence
for the Standards model’s notion that the standard is involved in the procedure of
metacomprehension assessments. Obviously, Ryan (1984) has demonstrated that readers’
epistemological beliefs affect their standards that are applied to evaluation of their reading
performance. Two other studies have provided somewhat direct evidence to support the impact
of epistemological beliefs on metacomprehension assessments, absolute accuracy in particular.
Schommer (1990) showed that the more naïve epistemological beliefs the reader holds the more
likely he or she overestimates. The study result implies that readers with a low standard are more
likely to make overestimations than those with a high standard. From the Standards model’s
stand point, one’s standard level has to do with his tendency to over- or under-estimate. To
illustrate, if one holds a low standard regarding what constitutes good comprehension, he is expected to display the tendency of exaggerating his reading performance. In contrast, if one applies a high standard when he evaluates his reading performance, he will underestimate his performance to a varying extent, depending on how high the standard is. Hence, it is said that the study results saying that naïve theorists are more likely to overestimate than sophisticated ones supports the Standards model. Schommer and colleagues (1992) have also found that sophisticated theorists tend to be more accurate when retrospectively evaluating their reading performance than naïve theorists. The finding can be interpreted that readers with a high standard are more capable of evaluating their test performance accurately than their low standard counterparts when making postdictions. Although the Standards model does not offer a specific explanation of how the level of standard affects the extent of accuracy, this study result as well as the previous one is, at least, said to support that the standard of good comprehension is involved in the procedure of the metacomprehension assessments.

**Accounting for Limited Absolute Accuracy of Predictions**

Remember that, from the Anchoring and Adjustment model’s view, one’s self-perception of reading skill, which is used as the anchoring point, is distorted from the actual level of reading skill. This distortion is thought to be the exclusive reason for low absolute accuracy, assuming a large dependency of metacomprehension predictions on the anchoring point. Comparably, the Standards model posits that limited absolute accuracy is attributable to the inappropriateness of the standard that individuals use as the reference point. If one uses an appropriate standard, the average of his predictions can be close to the mean actual performance even when their fluctuations do not correspond. Instead, use of an inappropriate standard gives one an illusion regarding the understanding level of the entire set of texts read in an experimental setting, which plays a critical role in lowering absolute accuracy. In this model, readers are assumed to hold a
varying idea of good comprehension, presumably dependent upon their epistemological beliefs. Thus, they are viewed to employ a variety of standards when predicting their performance over upcoming comprehension tests. More specifically, one can hold any level of standard between a good text-base and a high quality of situation model. That individual readers use a different standard implies that a large proportion of them possibly use an inappropriate one. According to the Standards model, this is why readers make predictions that are far from the actual test performance.

What is the appropriate level of standard, then, that readers should hold to make accurate predictions? Note that both absolute and relative accuracy are computed based on comparisons between predicted and actual performance in certain comprehension tests given in each study. Therefore, accuracy of predictions is largely affected by what the criterion tests require from readers (Wiley et al., 2005). In other words, the test features (e.g., type of testing items) can affect readers’ metacomprehension assessments (see also Maki, Willmon, & Pietan, 2009). However, Wiley and colleagues did not provide a concrete explanation of how and what kind of test features affects absolute accuracy in particular. Thus, I have attempted to delineate the impact of the criterion test on absolute accuracy from the Standard model’s view. Primarily, I note that there should be a person who devised the comprehension tests, and the test-giver developed it based on his idea of what readers should learn from texts. To put it differently, the test-giver’s idea of good comprehension determines the construction of the comprehension tests. For instance, if a test-giver holds a belief that good comprehension is constructing a text-base representation, he would include a majority of memory-based questions in his comprehension tests. On the contrary, if another test-giver aims at a situation model constructed as a result of text processing, his comprehension tests would mainly consist of inference-based questions.
Notably, one can have a similar or a different idea of good comprehension as the test-giver. I predict that, for someone with a similar idea, his predictions are expected to be close to the actual performance in the criterion tests across the given texts. He would predict his test performance in reference to the standard that is similar to what the actual test requires. In contrast, if a reader has a greatly different idea of good comprehension from the test-giver, his predictions will likely be far from the actual test performance on average due to the inappropriateness of the standard employed. Hence, from the Standards model’s perspective, how similar an idea of good comprehension one holds as the test-giver is critical in determining his absolute accuracy of predictions. Consequently, this model hypothesizes that the gap in the standard between individual readers and the test-giver explains their absolute accuracy.

There are two ways in which absolute accuracy is limited: one can either under- or over-estimate their reading performance. As explained previously, in general, a good situation model is more difficult to construct than a good text-base model (McNamara et al., 1996; Millis et al., 1998). Inference-based and memory-based questions can be answered based on the situation model or the text-base that one constructs after reading a text, respectively. Taken together, in the current study, I assume that inference-based questions are harder to correctly answer than memory-based questions. Accordingly, it is expected that readers who set a higher standard than the test-giver likely underestimate their test performance across the given texts overall. To illustrate, high standard readers suppose a somewhat difficult test consisting of a number of inference-based questions as a valid comprehension test and predict their test performance in reference to such a supposition. But, it turns out that the actual criterion test is easier than they had assumed, as it contains relatively less inference-based questions and more memory-based questions. Therefore, their predictions of the test performance tend to be lower than the actual
test performance on average. Meanwhile, the opposite is the case for readers with a lower standard than the test-giver. Their concept of good comprehension allows them to suppose a relatively easy comprehension test as a valid one. It makes them overestimate their performance in the actual criterion test, which often turns out more difficult than they thought it would. In conclusion, it is said that the test-giver’s standard is the appropriate level of standard that readers should use as the reference for accurate predictions. Unfortunately, people employ a different level of standard, which is frequently higher or lower than the proper standard. Therefore, from the Standards model’s view, whether one’s standard is higher or lower than the test-giver’s standard is viewed to predict whether he under- or overestimates his test performance.

Lastly, it is noteworthy that an individual can use a varying standard depending on the text features (Wiley et al., 2005). Thus, it is plausible that, given the texts with a variety of set of features, one’s predictions fluctuate due to the different standard applied to each text, not variation in the extent of text understanding. Bråten and colleagues (2008) have claimed that one applies a different profile of epistemological beliefs to texts with a varying topic familiarity. For instance, when reading informative texts (e.g., expository texts) on a novel topic relative to familiar ones, readers are more likely to value retrieving what the author conveys, assuming that the author is more knowledgeable than themselves, rather than focusing on constructing their own meaning based on interpretation of the text ideas. Thus, it is expected that a greater number of readers apply a low level of standard, construction of a text-base, with that type of texts. Meanwhile, according to Wiley et al. (2005), the structural complexity of text material has a large impact on the standard. Given a text wherein ideas are linearly described and their relations appear explicitly, readers will only need to focus on memorizing text ideas during reading and value establishment of a text-base as the result. Inversely, with a text implicitly conveying
numerous complicated relations, there may exist relatively great individual differences in the
standard ranging from pursuing a good text-base to aiming at a situation model as the reading
goal. Therefore, standard can vary by text features as well as by individual readers.

Note that a test-giver, as an individual, can also apply a different standard to a different
comprehension test, depending on the text features (e.g., topic familiarity, structural complexity).
Hence, it is plausible or appropriate that one develops comprehension tests with a varying
construction (e.g., a different number of inference-based versus memory-based questions) across
texts with a different profile of features. Given a set of texts with a variety of characteristics,
thus, one’s perceptions of text understanding are possibly complicated by fluctuation not only in
his or her own standard but also fluctuation in the test-giver’s. Consequently, features of the texts
should be equalized in order to prevent those complications from occurring and to attain
interpretable outcomes about whether one’s standard is involved in metacomprehension
assessments. In other words, with a set of texts with unified features, both the test-giver and
readers are expected to consistently apply one standard to the all of the texts given. Then,
predictions would fluctuate only by change in the extent of text understanding, while those
standards remain fixed. Therefore, one can conclude with relative ease whether the gap in the
standard between the test-giver and readers explains absolute accuracy of predictions. I also want
to remind that, even with a set of texts of unified features, there can still be individual variances
in the idea of good comprehension as there exist individual variances in personal epistemology
(e.g., Schraw, 2000), although the magnitude of the variance changes depending on the text
features (e.g., structural complexity). In the upcoming part, I will discuss how to explain
individual variations in the standard and how they affect absolute accuracy of predictions,
assuming the relevant features are unified across the given texts.
Accounting for Individual Differences in Absolute Accuracy of Predictions

Empirical data have reported that subjects display a reliable absolute accuracy across occasions (Schraw & Roedel, 1994; West & Stanovich, 1997). The next question asked is: who keeps under- versus overestimating their reading performance when making predictions? On the basis of the results of previous studies (e.g., Glover, 1989), it seems that reading skill is a critical variable in understanding the consistent pattern. That is, less-skilled readers are more likely overestimate than their more-skilled counterparts. Remember that the Anchoring and Adjustment model explains that people differ in the manner of perceiving their own reading skill as a function of the skill level: more -skilled readers have a fairly strict self-perception of reading skill relative to the reality while less-skilled readers tend to be overly generous in terms of perceiving their reading ability. Comparably, assuming that use of an inappropriate standard successfully accounts for limited absolute accuracy, the Standards model hypothesizes that the divergence in absolute accuracy of predictions between more - and less-skilled readers is attributable to difference in the level of the standard each group employs. That is, more-skilled readers likely set a higher standard than their less-skilled counterparts, which resulted in the former group holding a higher or slightly lower standard and the latter with a greatly lower standard than the test-giver’s in many of the experimental settings. Such differences between skill groups are viewed to cause individual differences in absolute accuracy of predictions.

To rephrase the hypothesis by using the definition of the level of the standard discussed previously, more -skilled readers are viewed to more likely consider construction of a situation model as good comprehension, whereas less-skilled readers tend to conceive establishment of a text-base as their reading purpose. Since this model assumes that a situation model is more difficult to construct than a text-base (McNamara et al., 1996; Millis et al., 1998), more-skilled readers suppose a more difficult test with a larger proportion of inference-based questions as a
good comprehension test than their less-skilled counterparts. In comparison to such a high or a low standard of good comprehension that each group holds, the criterion test often turns out to be too easy or too difficult to a varying degree. Therefore, according to the Standards model, more-skilled readers are less likely to overestimate than their less-skilled counterparts due to the relatively too high (or too low standard for the less-skilled ones) they tend to use when making metacomprehension predictions.

There is abundant empirical evidence to support the notion that more-skilled readers hold a higher standard than less-skilled ones even if they are somewhat indirect. In many studies across domains (e.g., grammar, tennis, social skill) (see Kruger & Dunning, 1999, for a review), less-skilled people tended to be more generous in evaluating their own ability or performance than their more-skilled counterparts. Krueger and Dunning (1999) have interpreted the body of research that more-skilled people have a good understanding of how high level of skills are required to attain good performance because they possess that level of skills. By contrast, less-skilled readers are more or less incapable of recognizing what is needed to attain good achievement due to lack of such competences. Accordingly, less-skilled people are more likely to inflate their performance than more-skilled ones because of their looser appreciation of what is required for good performance in many areas. This seems consistent with the Standards model’s argument that more-skilled readers have a stricter standard of good comprehension than less-skilled readers, which yields the individual difference in metacomprehension predictions.

Eidelman and Biernat (2007) have also reported that more-skilled people set a higher standard than their less-skilled counterparts, but from a different perspective. According to them, by doing so, high performers can make their competence more conspicuous and continue, thus, to maintain their high self-esteem. On the flip side, low performers likely set a comparably low
standard against which their performance is assessed. If such empirical outcomes are the case for
the domain of reading comprehension as well, it is predicted that more-skilled readers who have
accumulated successful reading experiences likely set a higher standard than their less-skilled
counterparts, as the Standards model hypothesizes.

Although many studies have reported that high performers tend to set a higher standard
than low performers across domains, as aforementioned (e.g., Eidelman & Biernat, 2007;
Krueger & Dunning, 1999), they do not supply enough evidence to support the Standards
model’s hypothesis as they did not study on reading comprehension. Moreover, the standards in
those studies were not conceptualized in the exact same way as the Standards model does. Thus,
in order to attain more direct evidence regarding the difference in the standard as a function of
reading skill, I have inspected the relationship between reading skill and personal epistemology.
It is noteworthy that the study results were perplexing. Some studies have shown that reading
skill is predictive of epistemological beliefs (Buehl & Alexander, 2005; Schommer et al., 1999;
Ryan, 1984; Schraw & Bruning, 1996; Schraw et al., 2002). In those studies, sophisticated
theorists displayed a higher reading performance than naïve theorists. As explained earlier,
sophisticated theorists are thought to be readers with a high standard who consider the
construction of a situation model as good comprehension. Adversely, naïve theorists correspond
to those with a low standard who pursue a text-base built as a desirable product of reading
comprehension. Accordingly, the study results are interpreted that more-skilled readers had a
higher standard than their less-skilled counterparts, which is consistent with the Standards
model’s hypothesis. As opposed to those studies supplying affirming evidence, there exist other
studies that negated the Standards model’s claim (Bråten, Strømsø, & Salmuelstuen, 2008;
Mason, Scirica, & Salvi, 2006; Schraw, 2000). In these studies, no consistent relationships were
found between reading skill and the type of epistemological beliefs. If this is the case, it cannot be said that more-skilled readers hold a higher standard than low-skilled ones, which opposes to the Standards model. Overall, it is still in question whether the difference in metacomprehension predictions between the two reading skill groups can be explained by difference in the standard that each group employs.

When explaining the affirmative results attained, it seems that prior knowledge and reading skill are interdependent. Many studies have shown that prior knowledge is critical in performance in reading comprehension tests in general (e.g., Chiesi, Spilich, & Voss, 1979; Means & Voss, 1985; McNamara et al., 1996). Also, good reading skills are thought to be imperative to accumulating knowledge on a variety of topics, as texts are a crucial source of information. Hence, it is inferred that more-skilled readers, good performers in comprehension tests, generally have a larger amount of background knowledge than less-skilled readers, poor test performers. Importantly, one of the most crucial resources needed to build a good situation model is background knowledge. Text ideas should be interpreted based on and integrated with background knowledge in that type of model (Kintsch, 1994). Therefore, one can greatly benefit from rich background knowledge in terms of drawing different types of knowledge-based inferences that are essential to building a good situation model. Taken together, a solid knowledge basis may enable more-skilled readers to construct a high quality of situation model based on the utilization. In contrast, lack of background knowledge could deprive less-skilled readers of the chance of building a good situation model. Remember that many studies across domains have reported that lack of capability keeps people not only from making a high achievement but also from appreciating the value (See Krueger & Dunning, 1999, for a review). If this is the case for the reading comprehension, due to their relatively affluent knowledge basis,
more-skilled readers value a situation model as their reading goal as well as are capable of building it. By contrast, because of lack of the critical component for situation-model building, less-skilled readers cannot help but stick to conveying text ideas as the author intends to and constructing the consistent value system regarding reading comprehension. If this explanation rightly fits in data, the Standards model is thought to capture the individual difference in prediction accuracy as a function of reading skill.

However, a theoretical account also exists for the evidence negating the Standards model’s hypothesis. Schraw (2000) has noted that text-processing consists of automatic and strategic processes (see also McNamara, et al., 1996; van den Broek, Rapp, & Kendeou, 2005). According to Schraw (2000), one’s beliefs about good comprehension can have a critical impact on how he strategically manages reading processes, while it does not affect the operations triggered in an automatic, unconscious fashion. Since both automatic and strategic activities combine to determine one’s reading achievement, one’s concept of good comprehension cannot be a reliable predictor of his reading achievement. To illustrate, if a more-skilled reader mainly owes the superiority to high automation of a variety of reading sub-processes, even including higher order knowledge-based inferences, he can build a high quality of situation model even when, to him, good comprehension means construction of a text-base. Also, it is plausible that a less-skilled reader consciously attempts to put considerable effort in generating knowledge-based inferences in order to attain his goal of building a situation model but such efforts do not bear fruit due to inefficiency in the automatic processes. If this account is the case, it cannot be said that one’s reading skill level is predictive of the standard that he holds as to reading comprehension. Accordingly, the Standards model would fail to cover the individual difference in prediction accuracy as a function of reading skill level.
Accounting for Superior Absolute Accuracy of Postdictions over Predictions

Last, it is time to discuss how the Standards model explains the superior accuracy of postdictions compared to predictions. Remember that the Anchoring and Adjustment model attributes the enhanced accuracy to an increase in the adjustment due to enhanced confidence about their perceptions of understanding level of the text, and a decrease in the reliance on distorted self-perception of reading skill. The Standards model views that the superior accuracy owes to change in the standard used from predictions to postdictions. As explained earlier, this model ascribes limited absolute accuracy to use of inappropriate standard. Therefore, if accuracy of postdictions is superior to accuracy of predictions, it should be understood as use of more appropriate standard for the former than the latter.

How can the standards for postdiction be more appropriate than the standards for predictions in favor of accurate assessments? The Standards model supposes that the standard used for predictions is based on readers’ own conception of what it means to comprehend well. Readers often hold a different standard from the optimal one, that is, the test-giver’s idea of good comprehension. In contrast, when making postdictions, they have had an appreciation of the test-giver’s standard by taking the criterion test. The test-giver’s standard can be inferred from the construction, that is, the level of difficulty and content of test items, of the criterion test. I claim that one critical factor affecting such inferences is the number of inference-based or memory-based questions included in the criterion test. For instance, if a criterion test mainly consists of inference-based questions, readers would conclude that the test-giver employed a high comprehension standard when developing the test. Such an appreciation of the test giver’s idea of good comprehension allows readers to realize that their original standard was inappropriate in assessing their performance on the criterion test. If one feels that his original standard was too high relative to the test-giver’s standard, he would try to use a lower one for postdictions. If one
conceives that he employed too low a standard to make accurate predictions, the modified standard would be formed at a higher point than the original one. As a result, if this account is the case, the standards that readers build after seeing test items would cluster around the test-giver’s standard, whereas their original standards spread out over a somewhat wide range. Accordingly, it is expected that individual variance in the modified standard is smaller than that in the original standard (Figure 1-2). In addition, the standard gap between individual readers and the test-giver would be smaller for postdictions than for predictions. Remember this model claims that the bigger the gap in the standard between an individual reader and the test-giver, the lower the absolute accuracy. Therefore, it is predicted that a decrease in the gap in the standard from predictions to postdictions is expected to explain increases in the absolute accuracy between the two occasions.
Figure 1-1. An example of high relative- but low absolute accuracy (left) and low relative- but high absolute accuracy (right)

Figure 1-2. Hypothesized differences in the individual variance in the standard employed for predictions and postdictions
CHAPTER 2
RESEARCH QUESTIONS AND HYPOTHESES

The primary objective of this study is to provide a systematic account for three typical phenomena surrounding absolute accuracy of metacomprehension assessments: (1) the absolute accuracy of predictions is typically quite low; (2) there exist individual differences in absolute accuracy of predictions as a function of reading skill; and (3) postdictions are more accurate than predictions. To achieve the stated objective, I tested two models, the Anchoring and Adjustment model (Zhao & Linderholm, 2008) and the Standards model (based on Wiley et al., 2005) to determine which model better explains those three different findings. Therefore, the specific research questions are as follows: First, which model explains the relatively low absolute accuracy of predictions? Second, which model explains individual differences in absolute accuracy of predictions between more- and less-skilled readers? Third, which model explains superior absolute accuracy of postdictions over predictions? I will provide detailed explanations of how to address each question below:

**Question 1: Which Model Explains Limited Absolute Accuracy?**

From the Anchoring and Adjustment model’s perspective, one critical source for limited absolute accuracy is readers’ inaccurate self-perception of reading skill (Zhao & Linderholm, 2008). To illustrate, according to the model, when making metacomprehension predictions, readers set an anchoring point around self-perception of reading skill, and make adjustments away from it only to a small extent. Consequently, the anchoring and adjustment procedure will result in predictions made at a point around their self-perception of reading skill, regardless of variance in the text understanding. Therefore, the model claims that self-perception of reading skill should be predictive of metacomprehension predictions. I also note that one’s objective level of reading skill is predictive of his or her actual reading performance on specific occasions.
Considering the definition of absolute accuracy, the mean difference between predicted and actual test performance, the more inaccurate one’s self-perception of reading skill is the lower the absolute accuracy. Therefore, I hypothesize that, if the Anchoring and Adjustment model is the case, the gap between objective level of reading skill and its subjective perception should predict absolute accuracy of predictions.

In contrast, according to the Standards model, low absolute accuracy is due to use of an inappropriate standard in the procedure of metacomprehension predictions. This model posits that individual readers have a varying idea of desirable reading comprehension from valuing a text-base representation of a text to aiming at a situation model as a reading goal. When predicting their performance on a comprehension test after reading a text, they use their idea of the definition of good comprehension as the standard against which the perceived extent of text understanding is evaluated. Consistently, this model supposes that, in each experimental setting of metacomprehension studies, the criterion comprehension tests are developed based on the test-giver’s idea of good comprehension. In other words, the construction of the comprehension tests (e.g., the number of memory-based versus inference-based questions) is determined by the test-giver’s comprehension standard. Therefore, I suspect that, if an individual holds a similar concept of good comprehension to the test-giver’s, his predictions would approximate the actual performance in the criterion tests on average. By contrast, if an individual has a completely different idea of good comprehension from the test-giver, he would have an illusion about his overall test performance and end up with under- or overestimations, because the standard applied is overly high or overly low, respectively. Therefore, I predict that the greater the gap in the standard between the test-giver and readers, the lower absolute accuracy. To conclude, my hypothesis is that the more an individual reader differs from the test-giver in terms of the extent
of pursuing a situation model relative to a text-base as good comprehension, the lower the absolute accuracy he would display, if the Standards model is correct.

On the basis of these two different hypotheses proposed from each model’s perspective, the current study compared these two types of discrepancies, the gap between subjectively perceived and objectively measured reading skill and the gap in the standard between the test-giver and readers, and showed which type of gap would predict absolute accuracy:

- Hypothesis 1-1: The gap between subjectively perceived and objectively measured reading skill predicts absolute accuracy (Anchoring and Adjustment model).
- Hypothesis 1-2: The gap in the standard between the test-giver and individual readers predicts absolute accuracy (Standards model).

**Question 2: Which Model Explains Individual Differences in Absolute Accuracy of Predictions?**

Previous studies have consistently reported that more-skilled readers are less likely to overestimate than their less-skilled counterparts (e.g., Glover, 1989). As explained previously, according to the Anchoring and Adjustment model, the extent to which self-perception of reading skill deviates from the objective reality should account for low absolute accuracy of predictions. Consistently, the model attributes the individual differences to the manner in which each group perceives their own reading skill. Specifically, it is hypothesized that more-skilled readers tend to have lower self-perceptions of their reading skill, whereas less-skilled readers have largely higher self-perceptions of their reading skill relative to its objective level. Such more or less strict self-perceptions of reading skill are viewed to account for more-skilled readers’ tendency to underestimate. In contrast, less-skilled readers’ somewhat overly generous self-perceptions of reading skill are said to support their overestimations in terms of metacomprehension predictions.
Comparably, the Standards model ascribes limited absolute accuracy of predictions to the use of a standard which differs from the test-giver’s standard. Assuming that that is true, it is expected that more-skilled readers frequently have a higher standard than the optimal one, which yields underestimations in general. Oppositely, less-skilled readers likely hold a lower standard than the test giver's, which renders a relatively large degree of underestimations. In the present study, the criterion tests were designed to contain the same number of memory- and inference-based questions, as many previous studies on metacomprehension accuracy did (e.g., Maki et al., 2005; Rawson, Dunlosky, & Theide, 2000). Assuming that how many questions a comprehension test contains for each type represents the test-giver’s idea of good comprehension, it implies that the test-giver’s standard was manipulated to center around the medium point on the spectrum between one end of pursuing a text-base and the other end of valuing a situation model as a reading purpose. Therefore, it is hypothesized that more-skilled readers’ standards are likely to be higher than the medium point, pursuing a situation model as a reading purpose, whereas less-skilled readers’ standards tend to be lower than that, valuing a text-base as a result of desirable comprehension.

In order to reveal which model’s account covers individual differences in the absolute accuracy of predictions as a function of reading skill, I tested the following two hypotheses constructed from each model’s perspective:

- **Hypothesis 2-1:** More-skilled readers likely have a lower self-perception of reading skill than it actually is, while less-skilled readers have a significantly higher self-perception of reading skill than its objective reality (Anchoring and Adjustment model).

- **Hypothesis 2-2:** More-skilled readers’ standards tend to be higher than the test-giver’s while less-skilled readers’ standards are significantly lower than the test-giver’s (Standards model).
Question 3: Which Model Explains Superior Absolute Accuracy of Postdictions over Predictions?

The Anchoring and Adjustment model posits that, when making postdictions as well as predictions, readers anchor around their self-perception of reading skill and make adjustments away from the anchoring point to the perceived extent of text understanding (Zhao & Linderholm, 2008). However, when making postdictions, they are more confident about their perceptions of the text understanding level. The increased confidence derives from a significant situational difference between the occasion for predictions and that for postdictions: whether readers took the criterion tests. The difference causes a decrease in test uncertainty and an increase in cue validity, which can improve absolute accuracy. Regarding the procedural difference, meanwhile, the enhanced confidence about their perceived text understanding is expected to increase the magnitude of adjustments for postdictions relative to predictions. Accordingly, it is expected that postdictions are less dependent on the anchoring point, that is, self-perception of reading skill, while more dependent on the cues available, than predictions.

In contrast, the Standards model proposes that, for both postdictions and predictions, readers set a standard and evaluate the perceived extent of text understanding against it. According to the model, however, there exists one notable difference for each type of assessment, which is the standard employed. When making predictions, people use their own idea of good comprehension as the standard. The idea of good comprehension varies across individuals and normally differs from the test-giver’s idea of good comprehension to some degree. By contrast, individuals employ a new standard for postdictions. Throughout taking the comprehension tests, they have realized what the test-giver requires them to learn from the texts and how inappropriate their original standard was in terms of accurately predicting their test performance. Remember that, in the present study, the criterion comprehension tests were
manipulated to consist of the same number of inference- and memory-based questions. Accordingly, the test-giver’s standard was presumed to be set somewhere around the medium point between pursuing only a text-base representation and valuing a situation model as a reading purpose. Therefore, when making predictions, individual readers whose original standard is set at a point above this level would form a lower standard than the original one, when making postdictions. In contrast, readers with a standard lower than this level would enhance their standard for postdictions compared to the one for predictions. Consequently, the new standards are expected to converge toward the test giver’s standard, while the original standards spread out over a relatively broad range. Thus, individual variance in the standard is expected to be smaller for postdictions than for predictions. Further, recall that the standard gap between the test-giver and individual readers are expected to account for absolute accuracy of predictions. Therefore, increased absolute accuracy for postdictions should be explained by the decreased gap in the standard between the test-giver and individual readers.

On the basis of these explanations from each of the two models, two hypotheses are constructed as follows:

- **Hypothesis 3-1:** (1) The magnitude of adjustments is greater for postdictions than for predictions. (2) The gap between subjective and objective reading skill is a better predictor of absolute accuracy of predictions than absolute accuracy of postdictions (Anchoring and Adjustment model).

- **Hypothesis 3-2:** (1) Individual variance in the modified standard (developed after seeing test items) is smaller than the individual variance in the original standard. (2) A decrease in the gap in the standard between individual readers and the test-giver explains the increase in absolute accuracy from predictions to postdictions (Standards model).
CHAPTER 3
RESEARCH METHODS

Participants

The experiment was carried out over two separate sessions, which will be explained in
detail later. A total of 129 students at a large southeastern university participated in the first
session of the experiment, and 103 of those students came back for the second session. However,
8 out of the 103 students were screened out because their data were incomplete. Therefore, data
analyses were limited to the 95 students who fulfilled both sessions of the experiment. Among
them, 68 students were recruited from introductory educational psychology courses and other
social science courses. They were rewarded by course credits and $10 for each of the two
sessions, respectively. The remaining 17 students participated in the study after seeing flyers
posted around the campus and individually contacting the researcher. Their participation was
compensated by two $10 payments.

The demographic information of the participants is as follows: 73 students (76.8%) were
female and 22 (23.2%) were male. They ranged in age from 18.67 to 31.33, ($M = 21.41$ years
old, $SD = 2.13$ years old). All of them were native speakers of English. When categorizing them
by race, 51 (53.7%) were Caucasian/White, 21 (22.1%) were Latino/Hispanic, 16 (16.8%) were
African American, and 7 (7.4%) were Asian/Pacific. Regarding participants’ majors, 12 students
(12.6%) majored in liberal arts (e.g., Spanish, History), 48 students (50.5%) were from social
sciences and the applied areas (e.g., psychology, journalism), and 28 students (29.5%) were from
natural sciences and the applied areas (e.g., biology, nursing). 7 students (7.4%) were in fields of
art (e.g., dance performance, graphic design) or reported that their study area had yet to be
decided.
Materials

Self-Perception of Reading Skill Task

This self-report inventory was administered to investigate how students perceive their own level of reading comprehension skill. The texts over which comprehension was predicted or postdicted were on science topics in an expository text format. Meanwhile, studies have reported that students have greater difficulty in comprehending expository texts with scientific topics than narrative texts (e.g., Berman & Nir-Sagiv, 2007; Singer, Harkness, & Stewart, 1997; Wolfe, 2005). Thus, it was assumed that students had a different profile of perceptions of their capability in comprehending science texts versus narrative texts even if no empirical evidence has been provided. Overall, the inventory was specifically designed to measure self-perception of reading skill involved in processing of science texts.

The following instructions were presented before the self-report items in order to guide the students to recall their experiences with science texts and comprehension tests covering them.

You have many years of experience reading a variety of texts on science topics (e.g., magazine articles on global warming, newspaper articles on game addiction, on-line text materials on dinosaur extinction, etc.). In your years as a student (middle school, high school, and now college), you have also taken a number of reading comprehension tests over a variety of science subjects (e.g., psychology, biology, sociology, physics etc.). You have also probably evaluated how well you understood a science text on your own after reading as well – especially if you are reading in preparation for an exam. Please reflect on those experiences reading science texts and taking reading comprehension tests as you answer the following questions.

This questionnaire consisted of six self-report questions asking students’ self-perception of their level of ability in regard to several reading-relevant sub-skills or factors (Figure 3-1). Students were asked to answer each of these questions on an 11-point Likert scale with 1 indicating “the lowest skill level” and 11 indicating “the highest skill level”, except for Question 6, which is worded in a negative way. The sum of the ratings for all six questions indicates the
participant’s self-perception of reading skill with the rating for Question 6 reverse-coded. The raw scores were transformed to standardized T scores in order to be compared to their scores on the Reading Skill Test, of which introduction follows immediately. This scale was found to have an inter-item reliability (Cronbach’s Alpha= .88), and to moderately correlate with the Reading Skill Test score, \( r(93) = .41, \ p < .01. \)

**Reading Skill Test**

A test was administered to estimate students’ objective level of reading skill. It consisted of four texts and ten multiple-choice questions for each text. Each question had four possible answers, one of which was the correct answer.

To explain the texts in detail, they were adapted from the Scientific American or Time magazines, but received some minor modifications so that ten proper comprehension questions could be devised based on the content (Appendix A). Each text dealt with a unique topic from a different area. The topics were “Is music good for you?,” “Lost? Evidence that sense of direction is innate,” “Black carbon: An overlooked climate factor,” and “Why do heavy drinkers outlive nondrinkers?” The texts ranged from 595 to 724 words and the Flesch–Kincaid readability scores from 11.9 to 14.9 so that all texts could be understood by the college students in the study sample (Klare, 1975). The text materials used for this study were similar to the ones used in previous studies on metacomprehension accuracy in several aspects including genre, length, and difficulty (e.g., Anderson & Thiede, 2008; Rawson, Dunlosky, & Thiede, 2000) (Table 3-1).

Even if they vary in terms of topic, all four of the texts fall into the same genre with the same structure: an expository text with a causal structure, according to the expository text classification by Meyer and Freedle (1984). Expository texts were chosen because most studies on metacomprehension have employed that kind of text as their stimuli (e.g., Glenberg et al., 1987; Maki et al., 2005; Rawson, Dunlosky, & Thiede, 2000). By using text materials of the
same genre, the results of the current study could be interpreted in the context of other, similar studies. Expository texts can also be sub-categorized into different types by the structural complexity such as the collection text, the comparison text, and the causal text (Meyer & Feedle, 1984). Among them, causal texts were adopted for this study because these texts’ ideas are interconnected in logical and causal associations, whereas, in the other two types of texts, idea units are linearly arranged or simply juxtaposed for the purpose of comparison or contrast. Wiley et al. (2005) pointed out that the former requires one to build a situation model by reorganizing text ideas explicitly presented and uncovering implicit relations among them for in-depth understanding. In contrast, with the latter types of texts, due to the structural simplicity, building a text-base is enough to reach a full understanding of them. Therefore, causal texts are said to be the appropriate type of text materials not only to evaluate one’s capability of comprehending text ideas (not just memorizing them), but also discriminate students with a high standard from ones with a low standard of good comprehension. Most of all, by unifying the structure as well as the genre of the texts, I attempted to diminish the possibility that participants apply a different standard to each text (Wiley et al., 2005). However, it was still plausible that students would apply a varying standard to each text due to varying familiarity of the topics (Bråten et al., 2008), which will be examined later in detail.

Ten multiple-choice questions with four answer choices were devised for each text as aforementioned (Appendix B). Questions were divided into two kinds. Five of them were memory-based questions with responses based on the memorization of more or less isolated facts explicitly present on text surface. The other five were inference-based questions that should be answered by making inferences, that is, connecting different pieces of text information from separate parts of the text or integrating them with relevant prior knowledge. With regard to
devising inference-based questions, caution was made to minimize the possibility that they could be answered solely based on prior knowledge without consulting the text concerned (Wiley et al., 2005).

One’s objective level of reading skill was indexed by the total number of correct answers chosen. Thus, the possible minimum raw score was zero and the maximum was 40. The raw scores were transformed to the standardized T-scores so that one’s objective level of reading skill could be compared to his or her self-perception of reading skill. The Spearman-Brown split-half reliability coefficient was .68 when the two of those texts and the corresponding questions were randomly allocated to one half, and the other two texts and their test questions were assigned to the other half. Students’ score on this test was found to be moderately predictive of their Verbal SAT score, $r(62) = .40, p < .01$.

**Comprehension Standards Task**

The purpose of this task was to measure individual readers’ original concept of good comprehension. The inventory included the same four texts and forty questions (half were memory-based questions and the other half were inference-based questions) as in the Reading Skill Test introduced above. For this task, however, participants were guided to imagine that they were developing a comprehension test covering those four texts. They were then asked to rate how appropriate each question was as a test item in evaluating one’s comprehension of the corresponding text on a 5-point Likert scale (1: very inappropriate, 5: very appropriate).

The comprehension standard was indexed by the ratio of the sum of his ratings for inference-based questions to the sum of his ratings for memory-based questions. That is, the higher ratings the reader put on inference-based questions relative to memory-based questions the larger the standard score was.
This inventory was found to have a fairly good inter-item reliability, Cronbach’s Alpha = .91 and .85 for each of twenty memory-based and inference-based questions, respectively. Additionally, note that the correlations in this comprehension standard were much stronger than the correlations in topic familiarity across the four texts (Table 3-2). According to Bråten et al. (2008), readers may apply a different standard of good comprehension to each text depending on topic familiarity. In order to test this familiarity hypothesis, prior to reading the four texts used for the Comprehension Standards Task, I asked participants to rate how familiar they were with each topic of the four texts on a 11-point Likert scale (e.g., How familiar are you with the topic “the effect of drinking on mortality rates”?). Whereas there were comparably strong correlations in the comprehension standard as shown above, weak correlations were found in topic familiarity across the texts. Moreover, the correlations in the comprehension standard did not fluctuate in accordance with the correlations in topic familiarity. Taken together, as opposed to what Bråten et al. (2008) suspected, it seems that readers apply a similar standard of good comprehension to each text regardless of topic familiarity.

For a validity check, I administered a different inventory to measure the same construct as the Comprehension Standards Task and examined the extent to which these two measures were correlated to each other. The new measure was a self-report inventory to investigate students’ theories or beliefs of what constitutes good comprehension of science texts. It consisted of six self-report questions. Three of them asked participants how important they thought memorizing text information was in comprehending science texts. The other three asked their beliefs regarding the impact of inference-making on comprehension of science texts (Figure 3-2). Participants were asked to answer all of the six questions on an 11-point Likert scale. The same scoring method applied to the comprehension standard was used to score this
questionnaire. That is, I computed the ratio of the summed ratings for the questions asking students’ beliefs on inference-making to the summed ratings for the questions asking their beliefs on memorization of text information. With regard to the reliability of this inventory, however, the inter-item consistencies were not high. Cronbach’s alpha = .47 and .67 for each of those memory-relevant and inference-relevant questions, respectively. Importantly, no significant correlation was found between this and the one captured by the Comprehension Standards Task \( r(93) = .03, p > .05 \). Therefore, I failed to confirm the construct validity of the Comprehension Standards Task.

This low correlation could be attributed either to limitations residing in the self-report inventory or to low validity of Comprehension Standards Task. I suspect that the former is the case for two reasons. Primarily, the data displayed a limited reliability with the new self-report inventory, while displaying a fairly good reliability with Comprehension Standards Task. Also, contrary to what was planned, it seems possible that the self-report inventory does not measure the same construct as the Comprehension Standards Task. For instance, note that the new questionnaire considers prior knowledge as a critical factor in making inferences (Question 5 in Figure 3-2), a notion that is supported theoretically and empirically (e.g., Kintsch, 1988; Mannes & Kintsch, 1987). However, readers can also benefit from rich background knowledge in regard to remembering text ideas (Kintsch, 1994; McNamara et al., 1996). Therefore, it was plausible that those with a low standard had a high score on the self-report inventory by rating the Question 5 significantly high. Consequently, even if the construct validity of the Comprehension Standards Task is still in question, it seems too hasty to conclude that it is invalid to measure one’s concept of good comprehension. Instead, I believe that there may exist other ways to validate it. Specifically, the comprehension standard can be conceptualized as a type of personal
epistemology as explained previously in the introduction section. Therefore, its comparison with epistemological beliefs will be one good way to validate it in future studies.

**Metacomprehension Accuracy Task**

This study used the typical method that many studies on metacomprehension accuracy have employed to measure participants’ metacomprehension accuracy. That is, immediately after reading each of the four texts given, participants were asked to proactively evaluate performance on an up-coming comprehension test covering the text. Specifically, they answered the question “How well do you think you will be able to answer multiple-choice test questions covering the text titled __________ in 20 minutes?” on percentile scales. After making predictions for all of those four texts, they were then instructed to answer six multiple-choice test questions for each text. Finally, they retrospectively estimated their performance in the test they had already taken by answering the question “How accurately do you think you have answered the test questions?” on percentile scales.

These texts were similar to those used in the Reading Skill Test and the Comprehension Standards Task in terms of the features including genre, structure, source, and difficulty level. However, these texts were a little shorter than the texts used for those two Tasks (Table 3-3).

The six multiple-choice questions were developed based on the same rules as the questions for the Reading Skill Test and the Comprehension Standards Task. Three of them were memory-based questions and the other three were inference-based questions. The performance in this criterion test was indexed by the total number of questions answered correctly. Thus, the plausible scores ranged from 0 to 24. The raw scores were transformed to scores on percentile scales so that they were comparable to predictions or postdictions made in percentages as well.

Note that the Standards model supposes that the standard of the person who developed the criterion test would be the appropriate level of reference for metacomprehension assessments,
and the gap in standard between individual readers and the test-giver would explain their low absolute accuracy. Therefore, in order to quantify the extent to which an individual’s standard deviated from the test-giver’s standard, it was necessary to calculate the test-giver’s standard as well as the individual’s standard. The individual’s standard was measured by using the Comprehension Standards Task, introduced above, and the same method was applied to quantify the test-giver’s standard. To explain, as seen in the previous paragraph, the current study manipulated the construct of the criterion test so that it contained the same number of memory-based and inference-based questions covering each text. Therefore, it was inferred that the test-giver put an equal weight on both types of questions from the construct of the criterion test. In other words, the test-giver demanded readers to memorize somewhat detailed text information and build an in-depth understanding of the text as a whole. Since the construct was assumed to be determined by the test-giver’s standard, the test-giver’s standard was thought to render the score of 1 because the sum of his ratings for inference-based questions and memory-based questions would be equal.

**Modified Comprehension Standards Task**

This measure aimed to investigate individual readers’ *modified* standard after taking the criterion test, whereas the Comprehension Standards Task was developed to measure the standard they *originally* hold before that. The same texts and the same questions were used for both tasks. Participants were also asked to rate those questions in terms of the appropriateness as a comprehension testing item on the same 5-point Likert scale, and the same method was used to score the modified comprehension standard as well as the original one. The only difference between the two tasks was that, in this task, participants were guided to recall the questions included in the criterion test of which performance they had predicted or postdicted, and get into the mind-set of the person who developed the test. By doing so, their ratings of the questions
were expected to reflect perception of the test-giver’s standard employed to develop the criterion test, not their own standard of good comprehension. The instructions were illustrated as follows:

There is a guy who developed the comprehension tests that you took today. He developed the 6 questions covering each of the 4 texts in the TASK V (which indicates the Metacomprehension Accuracy Task) based on his idea of the best comprehension test. Now, imagine that you are the test-giver and you are asked to create a comprehension test over the 4 texts (e.g., Is music good for you?) by using a subset of the 10 questions you were given during the Session 1. Which questions would you rate higher or lower than other questions? Please re-do the rating that you did previously but, this time, from the test-giver’s perspective that is inferred from the questions and the construct of the comprehension tests that you took today, NOT yours.

**Procedure**

All participants took all of the tests and tasks introduced previously (Table 3-4). The experiment was conducted over two occasions with about a week interval between them. This interval was chosen because it was considered enough to minimize plausible impacts that the tasks of the first occasion had on the tasks of the second occasion (e.g., being tired of performing several tasks). In the first session, they were tested on three variables including self-perception of reading skill, objective reading skill, and their original standard of good comprehension in that order. In the second session, they were measured on the self-report questionnaire designed to validate the Comprehension Standards Task, metacomprehension accuracy and, then, the standard supposedly modified after taking the criterion test.

**Session 1**

The Self-Perception of Reading Skill Task and the Reading Skill Test were computer-based tasks. Before taking these critical tasks, participants received three practice questions to learn how to manipulate designated keys and were guided to answer questions collecting their demographic information.
Upon the Self-Perception of Reading Skill Task, on the first screen, participants were presented with the instruction to remind themselves of their experiences with science texts and the comprehension tests previously taken. Then, the six self-report questions were presented one at a time on a separate screen with an 11-point Likert scale. Participants were instructed to answer the questions by using numbers keys and advance to the next screen by pressing the enter key when they were ready. As a part of this task, four questions were additionally administered to investigate how familiar participants were with each topic of the four texts used in the Reading Skill Test and the Comprehension Standards Task. The purpose of these questions was to check the reliability of the Comprehension Standards Task as explained previously.

The Self-Perception of Reading Skill Task was followed by the Reading Skill Test. The instructions were given on the first screen, stating that participants would be given 4 texts and 10 multiple-choice questions for each text. These instructions emphasized that they should carefully read the texts because they would not be allowed to reread the texts or consult them while answering questions. Each text was divided into 9 paragraphs and presented one paragraph at a time after the topic was presented. After reading the last paragraph, the 10 questions were also given one question at a time, and participants answered the question by using number keys. The corresponding text topic always appeared on the corner of the screen where all of the 9 paragraphs and the 10 questions were present. Participants were allowed to take as much time as they needed both for reading paragraphs and answering questions. The same procedure was repeated for the four texts and the questions, and the order of the text presentation was randomized for each participant.

After taking the Reading Skill Test, participants were instructed to ask the researcher for the paper materials for the next task, which was the Comprehension Standards Task, even if they
were blind to what it was for. Two separate materials were given. One of them was the primary booklet containing the same 40 comprehension questions as those in the Reading Skill Test, and the other one was the supplementary material including the same 4 texts that they just read in the Reading Skill Test. The instructions on the first page of the main booklet guided them to rate each question in terms of how appropriate it was in evaluating one’s comprehension of the corresponding text. On the next page, the topic of one of the four texts was presented at the top to remind participants of the text content. Under the topic, the corresponding 10 questions were presented with a 5-point Likert scale right below each question. The other three texts and the 10 questions for each text were presented in the same manner. Participants were allowed to take as much time as they needed and to consult the texts given separately. Four types of booklets were devised in which the texts were presented in a different sequence and participants were randomly given one of those booklets.

**Session 2**

Participants were guided to sit in front of a computer, and given a question on the first screen to make sure that they remembered how to manipulate the designated keys. They were requested to ask the researcher’s help if they did not. Then two questions for collecting demographic information were given on the subsequent two screens.

Before taking the two critical tasks, the Metacomprehension Accuracy Task and the Modified Comprehension Standards Task, participants were asked to answer 6 self-report items to inquire their idea of good comprehension, which was used to validate the Comprehension Standards Task. Participants responded to those questions one question at a time on an 11-point Likert scale by using number keys as they did in the previous session.

In the Metacomprehension Accuracy Task, participants were instructed to carefully read the four texts that would follow as they were going to take a comprehension test covering each
text. They were also informed that, after reading each text, they were asked to answer the query “How well do you think you will be able to answer multiple-choice questions covering the text titled “_____” in 20 minutes?” on a percentile scale. The way in which each text was presented was the exactly same as in the Reading Skill Test. They were not allowed to reread and look at the text when making the prediction on the scale with 0% (will answer none correctly), 10%, 20%, …,50% (will answer 50% correctly), 60%,……,100% (will answer all correctly).

After reading all of the four texts and making the four predictions, participants were given another set of instructions on a screen, saying that they would take a comprehension test consisting of 6 multiple-choice questions covering each text and, after taking the test, they would be asked to answer another query “How well do you think you have answered the multiple-choice questions covering the text titled “_____”? on a percentile scale. Participants answered the comprehension questions as they did in the Reading Skill Test and made a postdiction, which was prompted by the same query presented in the instruction page, on the scale with 0% (answered none correctly), 10%, 20%, …,50% (answered 50% correctly), 60%,……,100% (answered all correctly) for each text. Each participant was given the four texts in a random order and the 4 sets of the 6 questions in the same order as the texts were presented.

After completing the Metacomprehension Accuracy Task, participants were guided to ask for the material for the next task as they did in session 1. This time it was for the Modified Comprehension Standards Task. The same booklet was given with the same texts as they were given for the Comprehension Standards Task except for the instructions. For this task, they were instructed to get into the mind-set of the test-giver of the criterion test included in the Metacomprehension Accuracy Task they had taken immediately before. Participants rerated the questions from the test-giver’s perspective, which was inferred from the questions and the
construct of the comprehension test, in the exactly same way as the Comprehension Standards Task (Figure 3-3).
Table 3-1. Features of the texts used for the Reading Skill Test

<table>
<thead>
<tr>
<th></th>
<th>Topic</th>
<th>Source</th>
<th>Number of Words</th>
<th>Flesch-Kincaid grade level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text 1</td>
<td>Is music good for you?</td>
<td>Scientific American</td>
<td>595</td>
<td>14.9&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>Text 2</td>
<td>Lost? Evidence that sense of direction is innate</td>
<td>Scientific American</td>
<td>724</td>
<td>12.5&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>Text 3</td>
<td>Black carbon: An overlooked climate factor</td>
<td>Time</td>
<td>708</td>
<td>11.9&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>Text 4</td>
<td>Why do heavy drinkers outlive nondrinkers?</td>
<td>Time</td>
<td>701</td>
<td>12.7&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Table 3-2. Correlation coefficients of the comprehension standard (topic familiarity) across the four texts

<table>
<thead>
<tr>
<th></th>
<th>Text 1</th>
<th>Text 2</th>
<th>Text 3</th>
<th>Text 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text 1</td>
<td>.52* (.38*)</td>
<td></td>
<td>.65* (.18)</td>
<td>.61* (.32*)</td>
</tr>
<tr>
<td>Text 2</td>
<td></td>
<td>.63* (.23)</td>
<td></td>
<td>.65* (.16)</td>
</tr>
<tr>
<td>Text 3</td>
<td></td>
<td></td>
<td>.79* (.41*)</td>
<td></td>
</tr>
<tr>
<td>Text 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: *p < .01

Table 3-3. Features of texts used for the Metacomprehension Accuracy Task

<table>
<thead>
<tr>
<th></th>
<th>Topic</th>
<th>Source</th>
<th>Number of words</th>
<th>Flesch-Kincaid grade level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text 1</td>
<td>How did bacteria gobble methane from BP spill?</td>
<td>Scientific American</td>
<td>481</td>
<td>13.0&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>Text 2</td>
<td>How do we detect pancreatic cancer?</td>
<td>Scientific American</td>
<td>456</td>
<td>15.6&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>Text 3</td>
<td>How do we use facial expressions to convey emotions?</td>
<td>Scientific American</td>
<td>515</td>
<td>14.0&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>Text 4</td>
<td>What causes beak deformities in Northwest?</td>
<td>USA Today</td>
<td>455</td>
<td>13.4&lt;sup&gt;th&lt;/sup&gt;</td>
</tr>
<tr>
<td>Name</td>
<td>Purpose</td>
<td>Items</td>
<td>Score</td>
<td></td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>Self-Perception of Reading</td>
<td>To evaluate the level of reading skill that one subjectively perceives</td>
<td>6 self-report questions / 11-point Likert scale</td>
<td>The sum of the ratings for the 6 questions (Standardized)</td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>about him- or herself</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading Skill Test</td>
<td>To appraise the objective level of reading skill one displays relative</td>
<td>4 texts and 10 multiple-choice questions for each text (5 memory-</td>
<td>The sum of the number of correct answers out of the 40 questions</td>
<td></td>
</tr>
<tr>
<td></td>
<td>to others</td>
<td>based and 5 inference-based questions)</td>
<td>(Standardized)</td>
<td></td>
</tr>
<tr>
<td>Comprehension Standards Task</td>
<td>To assess one’s idea of good comprehension</td>
<td>The same texts and questions as the Reading Skill Test</td>
<td>The ratio of the summed ratings for inference-based questions to the</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>summed ratings for memory-based questions</td>
<td></td>
</tr>
<tr>
<td>Metacomprehension Accuracy</td>
<td>To assess how accurately one predicts or postdicts his or her</td>
<td>4 texts and 6 multiple-choice questions for each text/ a query to</td>
<td>The gap between predicted (or postdicted) performance (%) and actual</td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>performance in comprehension tests</td>
<td>prompt the prediction and one for the postdiction for each text</td>
<td>performance (%)</td>
<td></td>
</tr>
<tr>
<td>Modified Comprehension</td>
<td>To assess one’s appreciation of the test-giver’s idea of good</td>
<td>The same texts and questions as the Reading Skill Test</td>
<td>The ratio of the summed ratings for inference-based questions to the</td>
<td></td>
</tr>
<tr>
<td>Standards Task</td>
<td>comprehension</td>
<td></td>
<td>summed ratings for memory-based questions</td>
<td></td>
</tr>
<tr>
<td>The four topic familiarity</td>
<td>To evaluate one’s familiarity with the topics of the four texts in the</td>
<td>4 self-report questions /11-point Likert scale</td>
<td>Each represents one’s familiarity with each topic of the texts</td>
<td></td>
</tr>
<tr>
<td>questions</td>
<td>Comprehension Standards Task</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The reading theory self-report</td>
<td>To evaluate one’s idea of good comprehension</td>
<td>6 self-report questions (3 inference-relevant and 3 memory-relevant</td>
<td>The ratio of the summed ratings for inference-relevant questions to</td>
<td></td>
</tr>
<tr>
<td>inventory</td>
<td></td>
<td>questions) / 11-point Likert scale</td>
<td>the summed ratings for memory-relevant questions.</td>
<td></td>
</tr>
</tbody>
</table>

70
Q1. How well do you remember main ideas of a science text about 30 minutes after reading it?
Q2. How fast of a reader are you in terms of texts with a science topic?
Q3. How well do you integrate different pieces of text information (e.g., making cause-effect connections) while reading a science text?
Q4. How good of a reader are you in terms of texts with a science topic?
Q5. How well do you draw on background knowledge and use it to self-explain text ideas (e.g., guessing the meaning of unfamiliar words based on the relevant background knowledge) while reading a science text?
Q6. How much difficulty do you have retrieving main ideas of a science text when taking a comprehension test covering the content?

Figure 3-1. Six self-report questions for the Self-Perception of Reading Skill Task

**Memory-relevant questions**

Q1. How important role do you think the ability to recall text information plays in comprehending science texts?
Q2. How relevant do you think the difficulty of retrieving text information is to performance in a comprehension test covering a science text?
Q3. How much effort do you make to memorize what author meant during reading science texts?

**Inference-relevant questions**

Q4. How much impact do you think the ability to integrate text information during reading (e.g., making cause-effect connections, associating relevant text ideas, etc.) have on the result of comprehending science texts?
Q5. How much impact do you think the extent to which you have the background knowledge have on the result of comprehending a science text?
Q6. How much impact do you think the ability to interpret text ideas (e.g., coming up with examples relating to text information, evaluating text ideas, etc.) have on the results of comprehending science texts?

Figure 3-2. Six self-report questions used to validate the Comprehension Standards Task
Figure 3-3. The procedure in which participants took the tasks and tests

Session 1
- Self-Perception of Reading Skill Task
  (with the four topic familiarity questions)
- Reading Skill Test
- Comprehension Standards Task
  (About 1 week interval)

Session 2
- Reading Theory Self-Report Inventory
- Metacomprehension Accuracy Task
- Modified Comprehension Standards Task
Descriptive Statistics of the individual variables including the self-perception of reading skill, the objective level of reading skill, the original comprehension standard, and the modified standard are shown in Table 4-1. I computed standardized scores for the self-perception of reading skill and objective reading skill in order to compare these two scores to each other.

The Metacomprehension Accuracy Task generated three different scores including actual, predicted, and postdicted test performance. Actual test performance scores were converted into percentage scores so that they were comparable to the predicted and postdicted test performance scores made on a percentile scale (Table 4-2).

Absolute accuracy for both predictions and postdictions were computed based on those three scores (Table 4-2). As explained previously, absolute accuracy is indexed by signed discrepancy between the mean actual and the mean estimated test performance. The mean absolute accuracy was 6.47 (%) for predictions and 2.99 (%) for postdictions. Previous studies have reported various values of the mean absolute accuracy of assessments ranging from 1 to 15 (%) for predictions and 1 to 10 (%) for postdictions (Linderholm et al., 2008; Maki et al., 2005; Miesner & Maki, 2007). It can be said that the present study replicated the previous studies because both of the mean values are within each range of the existing data. However, these values do not provide information of the absolute accuracy that participants displayed on average. Data showed that many of them had a positive absolute accuracy with a varying absolute value while a considerable proportion had a negative absolute accuracy with a different absolute value. Thus, the sums of the individual values of absolute accuracy were close to zero. These mean scores were close to zero this much because of this reason, not because the individual absolute accuracy was actually that much high in general. In order to examine the
averaged absolute accuracy, thus, I took a look at their absolute values, regardless of the sign. The mean absolute value of absolute accuracy was 13.34% for predictions and 13.66% for postdictions. Thus, I concluded that, on average, participants’ predictions and postdictions differed from the actual performance by 13.34% and 13.66%, respectively.

To investigate the frequency data (Table 4-3), I divided participants into two groups in terms of the absolute value of absolute accuracy: those with a value below 10 (-10% < absolute accuracy < +10%) and those with a value above 10 (absolute accuracy < -10% or absolute accuracy > +10%). The criterion was selected because the standard deviation was around the score of 10, $SD = 10.37$ for predictions, $SD = 9.04$ for predictions. Regarding predictions, forty-three participants (45.3% of the entire sample) displayed absolute accuracy ranging between -10 and +10 percentage points, while fifty-two (54.7% of the entire sample) showed a level of absolute accuracy outside the range. Similar results were attained with regard to postdictions. Forty participants (42.1% of the entire sample) displayed an absolute accuracy of postdictions between -10 and +10, and fifty-five participants (57.9% of the entire sample) displayed an absolute accuracy outside the range. With the sign of absolute accuracy taken into consideration, twenty-nine (30.5% of the entire sample) held a negative value of absolute accuracy of predictions. That is, their predictions were lower than the actual performance (underestimation). Sixty-three participants (66.3% of the entire sample) had a positive value displaying higher predictions than their actual test performance on average (overestimation). When it comes to postdictions, thirty-four participants (35.8% of the entire sample) underestimated while fifty-nine (62.1% of the entire sample) overestimated their performance in the comprehension test that they had just taken when they retrospectively evaluated.
Which Model Explains Limited Absolute Accuracy of Predictions?

Data analyses showed that the Anchoring and Adjustment model appropriately explained the limited absolute accuracy. First of all, as assumed by the model, self-perception of reading skill reliably explained the predictions made in the Metacomprehension Accuracy Task, $\beta = .74$, $t(93) = 4.72$, $p < .01$, and the objective reading skill predicted the actual performance of reading comprehension over the four texts in same task, $\beta = .61$, $t(93) = 5.35$, $p < .01$. Most importantly, I computed the discrepancy between self-perceived and objective level of reading skill by subtracting the standardized T scores of the latter from that of the former. The discrepancy was significantly predictive of absolute accuracy, $\beta = .34$, $t(93) = 2.31$, $p = .02$, as the Anchoring and Adjustment model explained, even if the variance accounted for was small, $R^2 = .05$, $F(1,93) = 5.34$, $p = .02$.

Prior to testing the hypothesis built from the Standards model’s perspective, I checked that the data supported the assumptions underlying this model. The first assumption is that a situation model is generally more difficult to construct than a text-base. Accordingly, inference-based questions should be more difficult to correctly answer than memory-based questions. Data showed the assumption was met: on average, 72.27% of the participants correctly answered each of the twenty memory-based questions and 57.38% correctly answered the twenty inference-based questions in the Reading Skill Test. The same pattern was found in the criterion test used in the Metacomprehension Accuracy Task. In that test, 71.80% of participants chose the correct answer to each of the memory-based and 48.05% of the inference-based questions. Additionally, paired sample $t$-tests showed that participants attained a significantly higher performance on the memory-based questions than the inference-based questions both in the Reading Skill Test, $t(94) = 8.99$, $p < .01$, and the Metacomprehension Accuracy Task, $t(94) = 11.71$, $p < .01$. Thus, inference-based questions are said to be more difficult to answer. The data provides the basis for
the Standards model’s argument that the type of comprehension questions one prefers can affect whether they over- or underestimate their performance in upcoming comprehension tests.

Another important assumption was that the structure of the texts used in the Comprehension Standards Task was sufficiently complicated to discriminate those with a high standard from those with a low standard based on Wiley et al.’s (2005) suggestion. As a result of the data analysis, only a small proportion of participants were found to display a level of comprehension standard significantly higher than the score of 1, considered the value that a reader would hold when rating a text-base and a situation model equally as the product of good text comprehension. In contrast, a majority of them held a level of comprehension standard around or even lower than the score of 1 (Table 4-4). In other words, most participants valued building a text-base and a situation model similarly, or considered construction of a text-base a better way to comprehend a text than construction of a situation model. These results implied two different possibilities. One is that the structure of the texts used in the current study was not sufficiently complicated to discriminate those with a high standard of good comprehension from those with a low standard. Therefore, a certain proportion of students with a medium level of standard could apply a higher standard to texts with a greater structural complexity, assuming that the standard of good comprehension varies by the structural complexity of the text, as Wiley et al. (2005) proposed. The other possibility is that the data reflect the actual proportions of college students in terms of their concept of good comprehension at least with science or informational/expository texts that are relatively unfamiliar to them in terms of topic. That is, regardless of the structural complexity, whereas a small proportion consider building a situation model as a good comprehension, a large number of them value a text-base as their purpose of comprehending texts on a science topic.
When it comes to testing the Standards Model’s hypothesis that the gap in the standard between individual readers and the test-giver predicts absolute accuracy, regression analysis failed to confirm that the gap was a significant predictor of absolute accuracy $\beta = -4.53$, $t(93) = -.57$, $p = .57$. To explore the claims of the Standards model further, I attempted to divide participants into three groups in terms of how close their standard was to the test-giver’s standard: the high standard, the optimal standard, and the low standard group (Table 4-4). Those whose standard was beyond one standard deviation ($SD = .20$) above the test-giver’s standard (the score =1) were allocated to the high standard group. Those with a standard within the range one standard deviation above or below the test-giver’s were categorized into the optimal standard group. Those whose standard was one standard deviation or more below the test-giver’s belonged to the low standard group. According to the Standards model, the optimal group should display the highest level of absolute accuracy due to appropriateness of the standard that they employ. In order to test this hypothesis, these three groups were compared in terms of the absolute value of absolute accuracy, that is, the magnitude of the gap between predicted and actual test performance, regardless of the sign. As the result, even if the group difference was not significant, $F(2,92) = 2.02$, $p = .14$, as hypothesized, the optimal standard group tended to display a greater accuracy than the other two groups (Table 4-4). Additionally, I compared the three groups considering the sign of absolute accuracy. According to the Standards model, the low standard group was supposed to display the tendency of overestimations due to their overly low level of the standard applied. In contrast, the high standard group was expected to show the tendency of underestimations because the reference was set at a too high point to a varying degree. Data showed that the low standard group displayed the tendency of overestimation to the largest extent, which was followed by the optimal and, then, the high standard group (Table 4-4).
Thus, it can be said that, the standard level seems to affect one’s tendency of over-or underestimation the way that the Standards model supposes. Again, however, the group difference in absolute accuracy was not significant, $F(2,92) = .47$, $p = .63$. Consequently, even if the regression analysis failed to confirm the hypothesis that the gap in the comprehension standard between individual readers and the test-giver explains absolute accuracy, it seems too hasty to conclude that the Standards model’s hypothesis is untrue. Further investigation will be needed to clarify whether students use a standard when predicting their test performance after reading texts.

**Which Model Explains Individual Differences in Absolute Accuracy of Predictions?**

Prior to testing which model would explain this relationship between reading skill and absolute accuracy of predictions, I examined whether the same finding was obtained in the current study. Participants were divided into three groups according to their reading skill score: more-skilled, intermediatively-skilled, and less-skilled readers. Comparisons between the first and third group confirmed that the present study replicated the previous study’s results (e.g., Glover, 1989) in terms of the relationship between reading skill and absolute accuracy of predictions. That is, more-skilled readers overestimated less than their less-skilled counterparts on average, even if the accuracy difference was not statistically significant, $F(1,60) = 1.19$, $p = .28$.

Additionally, there were a greater proportion of readers who underestimated their performance among the more-skilled group (64.7%) than among the less-skilled group (35.3%). By contrast, among those who overestimated their test performance, more students were from the less-skilled group (56.8%) than the more-skilled group (43.2%). However, the group difference was not significant, $\chi^2(1) = 2.27$, $p = .13$, (Table 4-5).

According to the Anchoring and Adjustment model, this difference is attributable to more-skilled readers’ tendency of perceiving their reading skill lower than it actually is, and less-
skilled readers’ likelihood of holding a higher perception of their reading skill than the objective reality. In order to test the hypothesis, I conducted a split-plot analysis with the reading skill group as a between-subjects factor, and self-perceived and objective reading skill as a within-subjects factor. The analysis resulted in a significant interaction between the between- and the within-subject factor, $F(1,60) = 27.92, p < .01$ (Figure 4-1). Therefore, I have concluded that readers perceive their reading skill differently according to their reading skill level: more-skilled readers perceive that their reading skill is lower than the objective reality while less-skilled readers perceive that their reading skill is higher than it actually is. Such difference in self-perception of reading skill is said to explain the difference in absolute accuracy of predictions between the two reading skill groups, as the Anchoring and Adjustment model predicted.

From the perspective of the Standards model, this difference is caused by overly high or low standards that more- and less-skilled readers set relative to the test-giver’s, which is assumed to be the score of 1 in the current study. Even if, as shown previously, it is still in question that this model explains limited absolute accuracy, I attempted to investigate whether differences in the standard exist between more- and less-skilled readers. There was no significant difference in the standard between the two reading skill groups, $F(1,60) = 1.58, p = .21$. Thus, it is not the case that more-skilled readers have a higher standard than less-skilled readers. Accordingly, their strict or generous tendency of predictions cannot be explained by their difference in the standard set, opposing what the Standards model predicted.

**Which Model Explains Superior Absolute Accuracy of Postdictions over Predictions?**

As opposed to results attained from the general tendency of previous studies, a paired sample $t$-test resulted in no significant difference in the absolute value of absolute accuracy between predictions and postdictions, $t(94) = -.342, p = .73$. That is, the gap between postdicted and the actual performance did not significantly differ from the gap between predicted and the
actual performance. However, a previous study reported, depending on text difficulty, some
groups of participants show no accuracy difference in pre- and postdictions, and even greater
accuracy for predictions than postdictions (Maki et al., 2005). Thus, it cannot be said that the
current study result is without precedent even though it was unexpected. Additionally, even if the
accuracy did not differ between predictions and postdictions on average, a close examination of
each participant’s assessments over the four texts revealed that ninety-three (99%) out of ninety-five participants made a different set of postdictions from predictions across the four texts. The
data were interpreted to demonstrate a different psychological route taken for postdictions from
predictions. Thus, the exploratory attempt was still made to test each model’s hypothesis in terms
of what factors were involved in or what psychological procedures support postdictions versus
predictions.

According to the Anchoring and Adjustment model, the self-perception of reading skill is
applied as the anchoring point when postdictions as well as predictions are made. However, from
predictions to postdictions, it is expected that (1) the magnitude of adjustments increases, and,
accordingly, (2) dependency on self-perception of reading skill decreases. The data have
supported the first hypothesis formed from this model’s stance, that is, the magnitude of
adjustments increase due to readers’ enhanced confidence when making postdictions relative to
predictions. The inter-individual variance of the postdictions was significantly greater than that
of the predictions across the four texts given, $t(94) = 4.22, p < .01$. The result indicates that
participants made postdictions that fluctuated to a greater extent than predictions across the four
texts. Since the model assumes that each individual starts every assessment away from the same
anchoring point on one occasion, the amount of the variance in assessments (e.g., predictions)
can be interpreted to be the magnitude of the adjustments across the given texts. Thus, it seems
that participants made greater adjustments for postdictions than predictions, as the model hypothesized. In contrast, this model’s second hypothesis was not supported by the data. It was found that readers still use their self-perception of reading skill as the anchoring point when making postdictions as predicted. That is, self-perception of reading skill was found to be a good predictor for postdictions ($\beta = .89$, $t(93) = 5.76$, $p < .01$) as well as predictions ($\beta = .74$, $t(93) = 4.72$, $p < .01$). Surprisingly, however, regression analysis with self-perception of reading skill as the predictor yielded no significant difference in the $R^2$ square between when postdiction was the dependent variable ($R^2 = .26$, $F(1,93) = 33.22$, $p < .01$), and when prediction was the dependent variable ($R^2 = .19$, $F(1,93) = 22.29$, $p < .01$), $t(92) = 1.13$, $p = .13$. The gap between perceived and actual reading skill was even found to be a better predictor for accuracy of postdictions ($R^2 = .11$, $F(1,93) = 11.55$, $p < .01$, $\beta = .49$, $t(93) = 3.40$, $p < .01$) than that of predictions ($R^2 = .05$, $F(1,93) = 5.34$, $p = .02$, $\beta = .34$, $t(93) = 2.31$, $p = .02$), $t(92) = 1.70$, $p < .05$. The result was contradictory to what this model predicted.

Previous analyses have provided results likely opposing the Standards model as the valid account for the procedure of metacomprehension prediction. Since the model proposes that one takes the same route to both pre- and postdictions, except for change in the standard, it was suspicious that this model would be true about the procedure underlying postdiction. However, I attempted to explore whether there existed any difference between the original and modified comprehension standard. Primarily, it was expected that the modified standards would be closer to the test-giver’s standard than the original standards, due to participants’ appreciation of the test-giver’s idea of good comprehension by taking the criterion tests. On the basis of the expectation, I have hypothesized that individual variance in the modified standard is smaller than that in the original standard. However, data failed to confirm this hypothesis. There was no
significant change in the intra-individual variance between them, $SD = .20$ for the original comprehension standard and .24 for the modified comprehension standard (Table 4-1). The variance data indicated that participants’ modified standard did not converge into the test-giver’s standard even after taking the criterion test, which opposed what the Standards model predicted. The results imply that students did not appreciate the construction of the criterion test (e.g., the types and the number of the questions) as I hypothesized. Since this primary hypothesis was not supported, no analysis was carried out to test the second one, a decrease in the gap in the standard between test-giver and readers would account for the increase in absolute accuracy between predictions and postdictions.
Table 4-1. Descriptive statistics of individual variables

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-perception of reading skill*</td>
<td>95</td>
<td>23.58</td>
<td>70.77</td>
<td>50.00</td>
<td>10.00</td>
</tr>
<tr>
<td>Objective reading skill*</td>
<td>95</td>
<td>22.56</td>
<td>68.82</td>
<td>50.00</td>
<td>10.00</td>
</tr>
<tr>
<td>Original comprehension standard</td>
<td>95</td>
<td>.63</td>
<td>1.71</td>
<td>.94</td>
<td>.20</td>
</tr>
<tr>
<td>Modified comprehension standard</td>
<td>95</td>
<td>.59</td>
<td>2.10</td>
<td>.99</td>
<td>.24</td>
</tr>
</tbody>
</table>

Note: * Standardized T score

Table 4-2. Descriptive statistics of the Metacomprehension Accuracy Task

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual test performance (%)</td>
<td>95</td>
<td>20.83</td>
<td>95.83</td>
<td>59.96</td>
<td>12.67</td>
</tr>
<tr>
<td>Predicted test performance (%)</td>
<td>95</td>
<td>32.50</td>
<td>95.00</td>
<td>66.42</td>
<td>13.53</td>
</tr>
<tr>
<td>Postdicted test performance (%)</td>
<td>95</td>
<td>32.50</td>
<td>95.00</td>
<td>62.95</td>
<td>14.01</td>
</tr>
<tr>
<td>Prediction accuracy (%)</td>
<td>95</td>
<td>-29.17</td>
<td>56.67</td>
<td>6.47</td>
<td>15.66</td>
</tr>
<tr>
<td>Postdiction accuracy (%)</td>
<td>95</td>
<td>-38.33</td>
<td>44.17</td>
<td>2.99</td>
<td>16.16</td>
</tr>
<tr>
<td>Absolute value of prediction accuracy (%)</td>
<td>95</td>
<td>.00</td>
<td>56.67</td>
<td>13.34</td>
<td>10.37</td>
</tr>
<tr>
<td>Absolute value of postdiction accuracy (%)</td>
<td>95</td>
<td>.00</td>
<td>44.17</td>
<td>13.66</td>
<td>9.04</td>
</tr>
</tbody>
</table>

Table 4-3. Frequency data of absolute accuracy of predictions and postdictions

<table>
<thead>
<tr>
<th>Number of participants</th>
<th>Absolute value</th>
<th>Greater than 10* (%)</th>
<th>Smaller than 10 (%)</th>
<th>Underestimation (%)</th>
<th>Overestimation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictions</td>
<td>43 (45.3)</td>
<td>52 (54.7)</td>
<td>29 (30.5)</td>
<td>63 (66.3)</td>
<td></td>
</tr>
<tr>
<td>Postdictions</td>
<td>40 (42.1)</td>
<td>55 (57.9)</td>
<td>34 (35.8)</td>
<td>59 (62.1)</td>
<td></td>
</tr>
</tbody>
</table>

Note: * SD = 10.37 (%) for predictions, SD = 9.04 (%) for predictions

Table 4-4. Absolute accuracy of predictions according to the level of the comprehension standard

<table>
<thead>
<tr>
<th>Group</th>
<th>Standard level</th>
<th>N (% )</th>
<th>Mean of prediction accuracy (SD)</th>
<th>Mean of absolute value of prediction accuracy (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>~ .80</td>
<td>25 (26.32)</td>
<td>8.60 (17.90)</td>
<td>15.06 (12.69)</td>
</tr>
<tr>
<td>Optimal</td>
<td>.80 ~ 1.20</td>
<td>61 (64.21)</td>
<td>6.11 (13.62)</td>
<td>11.90 (8.90)</td>
</tr>
<tr>
<td>High</td>
<td>1.20 ~</td>
<td>9 (9.47)</td>
<td>2.96 (22.40)</td>
<td>18.33 (11.56)</td>
</tr>
</tbody>
</table>

* SD of comprehension standard = .20
Table 4-5. Differences in absolute accuracy of predictions between more-skilled and less-skilled readers

<table>
<thead>
<tr>
<th>Reading Skill Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>N. of participants who made underestimations</th>
<th>N. of participants who made overestimations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less-skilled readers</td>
<td>32</td>
<td>9.37</td>
<td>17.42</td>
<td>6</td>
<td>25</td>
</tr>
<tr>
<td>More-skilled readers</td>
<td>30</td>
<td>4.79</td>
<td>15.53</td>
<td>11</td>
<td>19</td>
</tr>
</tbody>
</table>
Figure 4-1. Difference in self-perception of reading skill between more-skilled and less-skilled readers
In this study, I attempted to explain three main common findings relevant to absolute accuracy of metacomprehension assessments: (1) the typically low absolute accuracy of metacomprehension predictions, (2) individual differences in absolute accuracy of predictions as a function of reading skill, and (3) the superior absolute accuracy of postdictions relative to predictions. Throughout the study, I compared two models: the Anchoring and Adjustment model (Zhao & Linderholm, 2008) and the Standards model (based on Wiley et al., 2005), and inspected which model better explained the three common findings comprehensively. This section includes a summary of the results and interpretations of findings, both expected and unexpected, for each model.

The Anchoring and Adjustment Model

The results of the present study suggest that, in general, the Anchoring and Adjustment model appropriately explains the three findings. To first summarize the study results, as the model claims, the gap between subjective and objective reading skill was found to be a significant predictor of absolute accuracy of predictions, although the gap accounted for a small variance in absolute accuracy. The result implies that other reasons exist for limited absolute accuracy as well. Second, data analyses showed that, for more-skilled readers, subjectively perceived reading skill was lower than the objective level, and the opposite was the case for less-skilled readers. Thus, it seemed that individual differences in the absolute accuracy of predictions were accounted for by the manner by which each group perceived their reading skill, as the model hypothesized. In regards to the third hypothesis, no significant difference was found in absolute accuracy between postdictions and predictions, which differed from the typical
results of previous studies (e.g., Hacker et al., 2000; Maki et al., 2005; Miesner & Maki, 2007). However, I still attempted to reveal the differences in the psychological routes taken because, for almost all participants, postdictions differed from predictions across the texts. Data analyses showed that, when making postdictions, readers relied on self-perception of reading skill to an even greater degree compared to when making predictions. This result opposed the model’s claims. By contrast, participants made greater magnitudes of adjustments for postdictions than predictions, which was congruent with the model. Overall, it can be said that the data generally fit into the Anchoring and Adjustment model, except for the one finding regarding postdictions’ greater dependency on self-perception of reading skill than predictions.

In order to interpret the one unexpected result regarding the increased predictive power of self-perception of reading skill for postdictions compared to predictions, it is needed to carefully reexamine how the hypothesis was formed and how it was tested. To address the former, the hypothesis that self-perception of reading skill would be a better predictor of predictions than postdictions was proposed because the magnitude of adjustments is viewed to increase for postdictions than for predictions, as the Anchoring and Adjustment model claims (Zhao & Linderholm, 2008). However, even if the hypothesis about the change in the magnitude of adjustments was confirmed, the hypothesis that self-perception of reading skill better predicts predictions than postdictions was not. The results imply that the Anchoring and Adjustment model has overlooked factors affecting the predictive power of self-perception of reading skill on predictions versus postdictions other than changes in the magnitude of adjustments. Note that, currently, the model assumes no difference in self-perception of reading skill employed for predictions versus postdictions (Zhao & Linderholm, 2008). However, I speculate that readers may perceive their reading skill differently because of a significant situational difference
between two the estimates: whether or not they have already taken the criterion comprehension tests. Specifically, when being asked to make predictions after reading texts, readers may be reminded of just their reading experience, not their experience with comprehension tests. They may use their self-perception of reading skill formed on that basis as the anchoring point. In contrast, when making postdictions, readers may self-perceive their own reading skill in a new way because the comprehension tests taken, as well as text-reading beforehand, allows them to more accurately perceive their reading ability. Therefore, it is plausible that participants in the current study employed a more accurate self-perception of reading skill as the anchoring point for postdictions than for predictions. Thus, the unsupported hypothesis may be understood better when considering these factors.

When it comes to how the hypothesis was tested, the current study developed a new self-report inventory, the Self-Perception of Reading Skill Task, to estimate one’s self-perception of reading skill, and inspected how well it accounted for predictions versus postdictions. Note that the Self-Perception of Reading Skill Task was different from the measures used in other studies on self-perception of skill across domains (see Dunning, Heath, & Suls, 2004, for a review). In previous studies, one simple question was normally employed to measure self-perception of skill in the domain (e.g., please rate your own social skills on the following 7-point Likert scale). By contrast, in the present study, the Self-Perception of Reading Skill Task asked how participants self-perceived their levels of different sub-skills involved in text processing (e.g., reading speed) and comprehension test-taking (e.g., memory ability) by using six different questions. Therefore, due to the concrete questions asking about reading sub-skills involved in both reading and comprehension test-taking, the Self-Perception of Reading Skill Task might have led participants to be more realistic about their own reading capability than simple questions adopted in other
Accordingly, the measured self-perception of reading skill plausibly reflected their objective level of reading skill with a greater accuracy in the present study than self-perceptions of different capabilities captured by relatively simple measures in other studies. Data confirmed this claim by displaying that, in the current study, the correlation between self-perceived and objectively measured reading skill was moderately high, \( r(93) = .41, p < .01 \). It contrasts to low correlation coefficients typically found between self-perception and one’s actual skills across domains other than reading (see Dunning, Heath, & Suls, 2004, for a review).

Taken together, it seems that my hypothesis overlooked two important points when predicting the smaller dependency of postdictions on self-perception of reading skill relative to predictions due to the increased magnitude of adjustments. One is that people can use a different self-perception of their reading skill as the anchoring point for the two occasions. Importantly, the anchors for postdictions seem more accurate than those for predictions because the criterion test-taking beforehand possibly make participants more realistic about their reading ability. The other is that the Self-Perception of Reading Skill Task might have induced participants to self-perceive their reading ability with a relatively high accuracy in the current study. Therefore, the measured self-perceptions of reading skill might have been more similar to self-perception of reading skills used for postdictions than those for predictions. Consequently, due to its greater similarity to the anchoring point adopted for postdictions, the predictive power of the measured self-perception of reading skill might have been greater for postdictions than for predictions, even with a greater magnitude of adjustments.

Empirical data appear to support the notion that the unexpected result can be explained by a greater similarity that the measured self-perceptions of reading skill have with the self-perceptions of reading skill used for postdictions as the anchoring point than they have with
those employed for predictions. In the present study, the self-perception of reading skill measured had a higher correlation with postdictions ($r(92) = .47, p < .01$) than prediction ($r(92) = .38, p < .01$), controlling for actual test performance. In contrast, in Zhao (2008), where a single, somewhat abstract question was used as the measure of self-perception of reading skill, the self-perception had a weaker correlation with postdictions ($r(57) = .26, p < .05$) than predictions ($r(57) = .35, p < .01$). The simple measure of self-perceptions of reading skill might only have captured vague, inaccurate self-perceptions of reading skill. Hence, maybe it was not the case that the measured self-perceptions of reading skill were more similar to the self-perception of reading skills used for postdictions than those for predictions in the study, assuming that the former were more accurate than the latter due to the situational difference (whether or not one had taken the criterion tests). Instead, a greater magnitude of adjustments could have a significantly decreasing impact of the dependency of postdictions on self-perception of reading skill. As a result, the estimated self-perception of reading skill accounted for a smaller variance of postdictions than predictions in Zhao (2008). In conclusion, the relative predictive power of self-perception of reading skill for postdictions versus predictions appears to rely on two factors at least: (1) a change in the magnitude of adjustments and (2) how similar the measured self-perceptions of reading skill are to those for postdictions versus for predictions, instead of solely depending on the adjustment magnitude, unlike the Anchoring and Adjustment model originally proposes.

In addition to the increased cue validity (Griffin, Jee, & Wiley, 2009), that one employs a more appropriate anchoring point based on a more accurate self-perception of reading skill for postdictions relative to predictions can be a new supportive account for the typical finding, superior absolute accuracy of postdictions relative to predictions (e.g., Hacker et al., 2000; Maki
et al., 2005; Miesner & Maki, 2007). By contrast, it makes the unexpected outcome attained in the present study, postdictions with no greater accuracy than predictions, more intriguing to interpret. In a search of appropriate interpretations, I noted that correlations in the actual performance in the criterion comprehension test were quite low across the four texts, .04 < r < .39. By contrast, correlations in the postdicted performance were comparably high across the four texts, .47 < r < .63. The difference in the magnitude of correlations could indicate that, even if fluctuation was larger for postdiction than for predictions, it was still not enough to fully reflect variation in the text difficulty. Variance data also confirmed this conjecture: variance in postdictions was significantly smaller than variance in actual performance across the four texts, t(94) = 6.32, p < .01. It was plausible that fluctuation in text difficulty was even greater for the current study than the previous ones. Hence, increase in the adjustments from predictions to postdictions might have fallen more short of sufficiently capturing fluctuation in text difficulty in the current study than in previous studies. This might be the reason that the current study generated postdictions only as accurate as predictions while postdictions were more accurate than predictions in previous ones.

Overall, the current study results support the Anchoring and Adjustment model as a valid account for metacomprehension predictions. That is, when making a prediction, college students anchor their estimate around their self-perception of reading skill and make an insufficient adjustment away from it. However, the model seems to be a little short of fully explaining the psychological route to postdictions. As the model posits, students were found to use their self-perception of reading skill as the anchoring point for postdictions as well as predictions, and the magnitude of adjustments were greater for the postdictions than for the predictions. However, the model appears to have overlooked that a different self-perception of reading skill is employed for
predictions versus postdictions. The self-perception of reading skill used for postdictions could be more accurate than that for predictions because taking the criterion tests beforehand could allow them to be more accurate about their reading skill than when making predictions only after reading the texts. Additionally, even when the magnitude of adjustments is larger for postdictions than for predictions, it is plausible that the enhancement is not enough to fully reflect fluctuation in text difficulty, which can result in postdictions with no more accuracy than predictions. To conclude, in order for the Anchoring and Adjustment model to comprehensively explain the mental procedure underlying both predictions and postdictions, abundant additional data will need to be collected on what changes take place in the anchoring point and to what extent the magnitude of adjustments increases from predictions to postdictions.

The Standards Model

It is time to interpret the study results in terms of how well the Standards model explains the three findings outlined at the beginning of this section. To first briefly summarize the results, unlike the model’s hypothesis on the limited absolute accuracy of predictions, a regression analysis failed to reveal the gap in the comprehension standard between individual readers and the test-giver as a significant predictor of absolute accuracy of predictions. However, when I divided participants into three groups by their own standard relative to the test-giver’s standard, and compared these groups in terms of the absolute value and the sign of absolute accuracy of predictions, the general pattern of the results were harmonious with the Standards model’s account, although not statistically significant. To be specific, the optimal group, of which standards were the closest to the test-giver’s standard, displayed the smallest discrepancy between predicted and actual test performance. Additionally, the low standard group, of which standards were lower than the test-giver’s, demonstrated the largest extent of overestimations. The extent was smaller for the optimal standard group, and it was the smallest for the high
standard group whose standards were higher than the test-giver’s. When it comes to the second hypothesis, there was no significant difference in the comprehension standard between more-skilled and less-skilled readers. The result opposed the model’s hypothesis that each reading skill group would be likely to hold a higher or lower standard than the test-giver’s, and such difference in the standard explains their tendency of under- or overestimation. Last, the modified standard of reading comprehension after taking the criterion tests did not converge around the test-giver’s standard as opposed to what the model suggested. Thus, it can be said that the model failed to explain how different the procedure for postdictions was from the one for predictions. Overall, the model explained the limited absolute accuracy at the very least, and did not account for the other two findings.

When it comes to the primary hypothesis on limited absolute accuracy of predictions, one plausible explanation is that it is just not the case that the comprehension standard is used as the reference for the metacomprehension predictions, unlike the Standards model’s supposition. That is, how much weight one puts on a situation model versus a text-base representation as the desirable result of reading comprehension does not affect one’s absolute accuracy. Instead, it is either possible that one uses a different type of standard or that no standard is involved in the procedure at all. However, I noted that the general pattern of the group difference data was congruent with the model’s account, as summarized above. Therefore, I attempted to explain why the present study did not support the hypothesis, assuming the results derived from some methodological problems residing in the present study, not in the model itself. As one possible limitation, I noted the finding that one of the two assumptions underlying the Standards model was violated. Remember that data confirmed the assumption that a situation model is more difficult to build than a text-base representation. By contrast, it was not supported that college
students’ idea of what consists of good comprehension varies a great deal, ranging from the one end of solely pursuing a text-base representation to the other end of aiming at a situation model as their reading goal. I suspect that this model’s failure in explaining limited accuracy of predictions might be relevant to this violation of the second assumption, that is, there was small variance in the comprehension standard across participants ($M = .94, SD = .20$). Specifically, only about 10% of participants held a standard higher than the mid-range standard (the score = 1), whereas about 65% of participants’ standard clustered around the mid-range standard, indicating that they pursued a text-base and a situation model to a similar extent. Note that the magnitude of the variance in the standard equals the magnitude of the variance in the gap in the standard between a reader and the test-giver, because the test-giver’s standard had a fixed value. Therefore, the comprehension standard’s lacking variance might be responsible for the result that the gap did not account for a significant amount of variance in absolute accuracy of predictions, not that the Standards model is incorrect about the procedure for metacomprehension predictions.

I conjecture that the small variance in participants’ ideas of good comprehension, that is, the comprehension standard, originated from the Comprehension Standards Task used to measure it. The measure might not have been sensitive enough to discriminate students with a high standard from those with a low standard. One possible reason might be the features of the texts (e.g., type, familiarity, structure) used as the stimuli in the Comprehension Standards Task. Note that the task employed expository texts on a science topic. The main purpose of expository texts is to provide factual information and to explain a more or less difficult mechanism (Wolfe, 2005). Thus, it might be the case that, when reading a science text in an expository format, a majority of people put their main focus on the construction of text-base rather than building a
situation model. Particularly, when students are unfamiliar with the text topic, they will spend time to just understand the author’s idea, and have no room to build their own beyond that (Bråten et al., 2008). I also consider the plausible failure in manipulation of the structural complexity as an important factor to decrease the magnitude of variance. Wiley et al. (2005) have suggested that people pursue building a situation model only when reading texts with a complicated structure. They do not need to do so otherwise because there is no significant difference between a text-base and a situation model built as a result of reading texts of which structure is relatively simple. Following the Wiley et al.’s suggestion, I attempted to choose texts with a sufficient structural complexity. But, if I failed, as opposed to the plan, the lack of structural complexity might have caused participants’ comprehension standards to have the small variance. Overall, as discussed until present, it seems possible that the level of comprehension standard that one applies varies depending on the features of the texts, and that the text features affect the magnitude of variance in the standard across individuals. Provided that they are true, thus, the gap in the comprehension standard between individual readers and the test-giver can explain absolute accuracy of predictions if a set of texts employed are appropriate enough to allow the measure to discriminate those with a high standard from those with a low standard. With a set of texts with the same profile of features, one needs to search for other ways (e.g., use a greater number of questions in the Comprehension Standards Task) to enhance the individual variation. However, this is only a hypothesis, and reaching any conclusion on this is beyond the capability with the results of the present study.

In addition to the primary hypothesis, it is also needed to take a close look at the finding that contradicted the second hypothesis: there was no significant difference between more-skilled and less-skilled readers on the comprehension standard. The result could also have derived from
the shortage of discriminatory power of the Comprehension Standards Task, not that the Standards model is theoretically incorrect. To reiterate, although it is the case that more-skilled readers hold a higher standard than their less-skilled counterparts as hypothesized, as in the current study, no significant difference could be found in the standard between the two reading skill groups because of the lack of variance in it. Particularly, if a set of texts with different features (e.g., texts with a higher structural complexity) had been given in the Comprehension Standards Task, or if it had included a greater number of questions, more-skilled readers could have displayed a significantly higher standard than their less-skilled counterparts. It would have allowed participants to be distributed across a wider range of the standard than they were in the present study. However, it is still possible that the hypothesis itself is wrong regardless of the text features. As explained previously, studies have provided perplexing results in regard to the relationship between reading skill and personal epistemology (e.g., Buehl & Alexander, 2005; Schraw, 2000). Provided that one’s idea of good comprehension, which is captured by the comprehension standard, reflects his or her personal epistemology, there is a fair amount of possibility that one’s reading skill level cannot predict his or her standard and vice versa. It may be because one’s epistemological beliefs can exert an impact on only strategically operated activities, not automatic ones during reading (Schraw, 2000). Consequently, further study is needed to reveal whether the unpredicted relationship found between reading skill and the comprehension standard is attributable to features of the Comprehension Standard Task, or because the hypothesis itself is incorrect.

Regarding the last hypothesis: according to the Standards model, when making postdictions, people use a modified standard that is different from the original one employed when making predictions. Readers are viewed to have the chance to appreciate what the test-
giver requires them to learn from the texts while taking the comprehension tests. Thus, it was hypothesized that participants would use a new standard that is similar to the test-giver’s, but the present study failed to verify it. Possibly, the hypothesis itself is wrong. That is, a change in the standard cannot explain the difference in the mechanisms used to make predictions versus postdictions. Since the model may not explain the mental route to metacomprehension predictions as found in the present study, it would not be the case either that the model covers the psychological procedure underlying postdictions. As another possibility, the model may have been incorrect about the way the change occurs in the standard from predictions to postdictions, whereas it is still true that there is some modification in it. To explain, the current study supposed that the test-giver’s standard would be determined by the number of inference-based versus memory-based questions in the criterion tests. The test-giver was supposed to hold the standard score of 1, assuming that the fact that the test included the same number of each type of questions indicated that the test-giver equally valued a text-base and a situation model as the reading standard. Likewise, I predicted that the number of each type of question would work as an outstanding factor to affect the participants’ appreciation of the test-giver’s standard. However, it was plausible that there were a variety of factors (e.g., difficulty of the questions) conspicuous to each of those participants. Those factors might have had a varying influence on individual participants’ impression of the test-giver’s idea of good comprehension. I speculate that it might have yielded a large variance in the modified standard, instead of a small one around the test-giver’s standard. In sum, this disconfirmed hypothesis could be another piece of evidence against the Standards model as the mechanism underlying metacomprehension assessments. However, if the model took factors other than the number of different types of test
questions into consideration, change in the standard could possibly have accounted for difference in predictions and postdictions.

Overall, the present study likely provided evidence that negated the Standards model rather than confirmed it. Instead, data have showed that metacomprehension assessments, including both predictions and postdictions, are reliably predicted by self-perception of reading skill as the Anchoring and Adjustment model suggested. Additionally, differences in self-perception of reading skill significantly explained individual differences in absolute accuracy of predictions between more-skilled and less-skilled readers. On the basis of the current study results, therefore, it is reasonable to conclude that the Anchoring and Adjustment model best explains psychological procedure underlying metacomprehension assessments, not the Standards model. However, it seems a little hasty to make a firm conclusion about the Standards model. It seems worthy to closely investigate ways to improve the discriminatory power of the Comprehension Standards Task to clarify whether the unexpected findings regarding the absolute accuracy of predictions derived from the small variance in it. Additionally, it is plausible that the model only overlooked different factors affecting one’s appreciation of the test-giver’s standard, not that the notion itself is untrue that people change their standard from postdictions to predictions.

Considering that the present study is pioneering work in exploring whether a sort of standard is involved in metacomprehension assessments, a great number of studies are needed to obtain solid evidence for or against it.

**Implications and Limitations**

This study results have several important theoretical and practical implications. Primarily the present study has provided an explanation for why absolute accuracy of predictions is relatively limited. Due to academic efforts focused on relative accuracy, it is known that low relative accuracy is attributable to invalidity of cues used as the basis for the perceived extent of
text understanding. However, attempts have rarely been made to reveal what causes low absolute accuracy, even though the weak correlation and the different patterns in reliability between the two types of accuracy have implied that an exclusive account exists for absolute accuracy separate from relative accuracy. The current study provides a partial account for absolute accuracy, although the results cannot be interpreted to offer the direct answer to the question, “What is the cause for limited absolute accuracy?” in a strict sense. I compared two different hypotheses: whether the limited accuracy is predicted by inaccurate self-perception of reading skill or by inappropriate standards applied. The result favored the former over the latter. That is, due to the large dependency of predictions on self-perception of reading skill, limited absolute accuracy can be explained by the extent to which one’s self-perception of reading skill is deviated from the objective reality. In sum, this study sheds light on the constraint for absolute accuracy of metacomprehension predictions.

The present study provides evidence to show the entire procedure that readers undergo when making metacomprehension predictions. Most studies have only demonstrated that the extent to which specific texts are comprehended is perceived based on cues available during the procedure (e.g., Koriat, 1997). The current study demonstrated that the behavior of anchoring takes place based on self-perception of reading skill, in addition to the adjustment-making process based on the perceived extent of text understanding. Zhao (2008) was the first empirical study that offered supporting evidence for the Anchoring and Adjustment model. In her study, predictions had a greater correlation with self-perception of reading skill, controlling for the actual performance. Additionally, the study showed that predictions changed with text difficulty to a lesser extent than the actual performance. Based on the results, she concluded that both anchoring and adjustments occurred in the procedure, and the anchoring had a diminishing
impact on the magnitude of adjustments. The present study confirms the model by adding supplemental, even stronger empirical evidence to Zhao’s. I tried to explain how the model covers major relevant findings including low absolute accuracy of predictions, the relationship between absolute accuracy of predictions and reading skill, and superior absolute accuracy of postdictions to predictions. Notably, the fact that data fit into most of the hypothesized accounts strongly supports the Anchoring and Adjustment model. In conclusion, through this study, I have attained a good starting point to understand the full mechanism underlying metacomprehension predictions.

Besides these major, intended implications discussed above, the current study has some additional but attention-deserving results with unique implications. First of all, it contributes to improving our knowledge about one’s theories or beliefs about what entails good comprehension, an important reading-relevant individual factor. Although the attempt was not successful in revealing how this factor was involved in the metacomprehension assessments, I have attained some notable information on college students’ concept of good comprehension specifically applied to science texts. Primarily, it appears that one pursues a text-base or a situation-model to a reliable extent as a reading goal across different texts, given they are equalized in terms of the genre and the structure. Next, upon reading science texts with a certain profile of features (e.g., structural complexity), a large proportion of college students appear to put a greater weight on construction of a text-base over situation model-building or a similar weight on both, whereas a relatively small proportion seem to aim at a situation model as their reading goal. Last but not least, the present study has provided a new tool to measure one’s epistemological beliefs specifically applied to science text comprehension, even if it needs to be verified by future studies. Existing inventories of personal epistemology uses somewhat
straightforward questions to investigate one’s ideas of the nature of good comprehension (e.g., To what extent do you agree or disagree with the statement “When I read a science text, I try to carry away exactly what the author meant”?). By contrast, this new tool takes an indirect approach to measure participants’ ideas of good comprehension (e.g., How appropriate do you think the following questions is as a comprehension testing item?). Thus, it is expected that the new tool is less likely to cause the validity problems (e.g., participants respond based on their thought of what is desirable to the experimenter, not of what is true about themselves.) than the existing ones. There are many questions remaining to be answered (e.g., what concept of good comprehension do college students hold with other types of texts?). However, the present study suggests a new perspective regarding investigations of college students’ concepts of good comprehension.

The current study provided critical information about how to measure adult readers’ self-perception of reading skill. Even if many studies have explored children’s self-conception of reading skill (e.g., Chapman & Tumner, 1995), few have been carried out on how adult readers perceive their reading capability. Adult readers’ self-perceptions of reading skill are supposed to be more sophisticated than children’s. Accumulated reading experiences appear to enhance their knowledge of text processing. Moreover, metacognitive abilities develop so that one grows more aware of his or her own capability with age (Schneider & Lockl, 2002, for a review). Thus, distinct measuring instruments are needed to inspect children and adult readers’ self-perception of reading skill. In the questionnaires investigating children’s self-perception of reading skill (e.g., Bouffard et al., 2003; Chapman & Tumner, 1995; Henk & Melnick, 1995), many of the items ask children to rate how great their general reading performance is (e.g., I read better than other kids in my class). Even if some ask on their perceived level of specific sub-skills, the
contents are restricted to relatively simple word-level skills (e.g., Are you good at remembering words?). In contrast, the Self-Perception of Reading Skill Task in the current study collected information on adult readers’ perceptions regarding high-level reading sub-skills (e.g., the ability of integrating text ideas) and factors impacting reading performance (e.g., reading speed) in a relatively detailed manner. Moreover, the task asks about one’s self-perception of reading-relevant skills that are particularly involved in processing science texts, assuming that such perceptions can vary by the text genre. The instrument was found to have an inter-item reliability and construct validity. Consequently, the present study is an initial step of investigation regarding how to measure adult readers’ self-perception of reading skill even if more precise research is needed to identify the full construct of it.

On the basis of the study outcomes, I propose a mechanism by which self-perceptions of reading skill affects reading achievement. It has been presumed that positive self-perceptions of reading skill have an enhancing impact on reading performance over negative ones (e.g., Henk & Melnick, 1995). One suggested mechanism is that positive self-perceptions of reading skill contribute in forming a positive attitude towards the behavior of reading. Such an affirmative orientation motivates readers to read more frequently, which, in turn, leads to good reading performance (Chapman, Tunmer, & Prochnow, 2000). However, some studies have shown that children’s attitudes towards reading or the relevant motivational profiles had only moderate correlations with reading achievement (Petscher, 2010, for a review). Additionally, correlations were not strong between self-perceptions and reading performance in previous studies on children (Cloer & Ross, 1996; Stringer & Health, 2008). Thus, I cannot simply claim that positive self-perceptions bring forth a high performance while negative ones lower it.
Note that the present study has showed that more-skilled readers are relatively strict while less-skilled readers are overly generous in evaluating their own reading capability (see also Kruger & Dunning, 1999, for other domains). I suspect that these different propensities in self-perceiving reading skill seem to affect reading achievements via metacomprehension predictions that they make on specific occasions. To illustrate, for more-skilled readers, their conservative self-perceptions may make them pessimistic towards their specific reading performance too, as showed in the present study. Importantly, this pessimism may lead them to devote even a greater time and effort to improve text understanding than they actually need. Conversely, due to the generous perception of their own reading skill, less-skilled readers are likely to be too optimistic towards their reading performance. They are thus expected to invest a lesser amount of mental resources than required to reach a full understanding. Accumulation of such occasions would help more-skilled readers continuously attain a high achievement, whereas less-skilled readers continue to have a low achievement in reading performance. Taken together, even if the subjects were adult readers in the present study while they were children in the previous ones, the current study result shows the possibility of an additional mechanism underlying the impact of self-perception of reading skill on reading performance. As previous studies claimed (e.g., Chapman, Tunmer, & Prochnow, 2000), positive self-perception can have an enhancing impact on reading achievement by enhancing motivation towards the behavior of reading. Comparably, the present study result implies that excessively positive self-perception can have a decreasing impact on reading achievements, and the opposite is the case for overly pessimistic self-perception of reading skill. These two mechanisms may combine to yield the relatively limited correlations between the two. In conclusion, this study contributes to improving our understanding of the relationship between self-perception of reading skill and reading performance.
In a nutshell, the current study has five major theoretical implications, as discussed above. First of all, it shows what constrains absolute accuracy independent from relative accuracy. Second, it contributes to demonstrating the entire procedure that readers undergo when making metacomprehension assessments. Third, the study results improve our knowledge of college students’ ideas or beliefs regarding what entails good comprehension of science texts. Fourth, it supplies good evidence regarding how to measure adult readers’ self-perception of reading skill. Last, it suggests a seemingly important link between self-perception of reading skill and reading achievement. Overall, by demonstrating that the Anchoring and Adjustment model is the case in terms of the procedure underlying metacomprehension assessments, the present study suggests that one’s self-perception of reading skill possibly impacts reading performance mediated by metacomprehension predictions on specific occasions. This finding could have meaningful educational implications, even if it is too hasty to claim that the relationship is causal based on the present study results. For example, students, less-skilled ones in particular, may need to be educated so that they are able to accurately perceive their own reading ability. I found that, they were overly optimistic about their reading ability, and tend to overestimate their specific reading performance. If such optimism is a cause for their low reading achievements as the present study suggests, helping them to have a realistic view of their reading ability can have an enhancing impact. One way of enhancing the accuracy of self-perception is to teach students what sub-skills are involved in text processing and what factors affect metacomprehension as suggested in the present study. Such efforts can allow them to become more precise in perceiving how good of a reader they are, to make more accurate assessments over specific reading performances, and, further, to make more efforts to improve their own reading performance.
However, one should be cautious when interpreting the current study results. Primarily, the reliability and the validity of the Reading Skill Test were not sufficiently high. Note that the score was used for two different purposes including measuring how inaccurately participants perceive their reading skill level and dividing them into two groups: more-skilled and less-skilled readers. Even if data confirmed the Anchoring and Adjustment model’s primary hypothesis, that is, inaccurate self-perception of reading skill significantly explained absolute accuracy of predictions, the explained variance was found to be very small. I suspect that, if a Reading Skill Test with a greater reliability (or validity) had been used, the explained variance would have been larger when considering that the hypothesis regarding individual differences between the two reading skill groups was supported.

Next, interpretations should be restricted to metacomprehension assessments in which science texts in an expository format are involved. All of the stimuli used to measure not only metacomprehension accuracy but also individual variables (e.g., self-perception of reading skill, one’s idea of good comprehension) were that type of texts. Text processing differs depending on the text genre (e.g., expository texts versus narrative texts) (e.g., Berman & Nir-Sagiv, 200). Accordingly, across different types of texts, adult readers can have a different idea of their own reading ability (e.g., Singer, Harkness, & Stewart, 1997; Wolfe, 2005) and of what constitutes good comprehension (e.g., Bråten et al., 2008; van den Broek et al., 1996). For these reasons, it is also plausible that one takes a different route when self-evaluating their reading performance given a different genre of texts.

Another important limitation is that this is a quasi-experimental design study. It means that, on the basis of the study results, I can only claim that college students’ self-perception of reading skill and their absolute accuracy are correlated. Thus, the possibility cannot be excluded
that the relationship is merely mediated by a third compounding factor. Moreover, as explained previously, the possibility cannot be ignored that the Standards model is the correct account for the mental route to metacomprehension assessments, although the data favored the Anchoring and Adjustment model. I suspect that the way readers perceive their reading skill level can have to do with the comprehension standard that they hold. For instance, those who have a too high or a too low standard of good comprehension may be incapable of accurately perceiving their level of reading skill. Thus, it may be the case that the comprehension standard is the third factor mediating the relationship between self-perception of reading skill and metacomprehension assessments. The current study supplies fairly convincing evidence that reading-relevant personal traits (e.g., self-perception of reading skill, one’s theory of good comprehension) is involved in the procedure of metacomprehension assessments. According to the study results, it seems more plausible that the trait would be self-perception of reading skill than personal theory of good comprehension. In order to confirm which one is the case, however, future studies should experimentally control one of them, and, then, observe what impact the other has on the psychological procedure of metacomprehension assessments.

Even with such limitations, the present study can still inspire not only subsequent studies on metacomprehension assessments but also studies on related topics such as the measurement of personal epistemology. Further, I hope that the study results will contribute to improving reading instructions, by informing readers more about what goes into good comprehension and what factors potentially bias readers’ estimates of comprehension.
APPENDIX A
A SAMPLE TEXT FOR THE READING SKILL TEST

Why Do Heavy Drinkers Outlive Nondrinkers?

One of the most contentious issues in the vast literature about alcohol consumption has been the consistent finding that those who don't drink tend to die sooner than those who do. The standard Alcoholics Anonymous explanation for this finding is that many of those who show up as abstainers in such research are actually former hard-core drinkers who had already incurred health problems associated with drinking.

But a new paper in the journal Alcoholism: Clinical and Experimental Research suggests that — for reasons that aren't entirely clear — abstaining from alcohol does tend to increase one's risk of dying, even when you exclude former problem drinkers. The most shocking part? Abstainers' mortality rates are higher than those of heavy drinkers.

Moderate drinking, which is defined as three drinks a day for men and one drink a day for women, is associated with the lowest mortality rates in alcohol studies. Moderate alcohol use, especially when the beverage of choice is red wine, is thought to improve circulation and respiration. Studies have also reported that daily consumption of a small amount of pure alcohol can slow the onset of diabetes. Additionally, enjoying a bit of alcohol seems to have an enhancing impact on their sociability, which can help to reduce mortality rates; people who are isolated don't have as many family members and friends who can notice and help treat health problems.

But why would abstaining from alcohol lead to a shorter life? Even if the exact reason hasn't been discovered, one of the most plausible suspects is high stress level. It's true that those who abstain from alcohol tend to be from lower socioeconomic classes, since drinking can be expensive. And people of lower socioeconomic status have more life stressors — job and child-care worries that might not only keep them from the bottle but also cause stress-related illnesses over long periods. They also don't get the stress-reducing benefits of a drink or two after work. Aside from high stress level, as the explanatory factor, scientists have suspected level of physical activity, dietary habits, and accessibility to medical services, which have something to do with socioeconomic status as well. But even after controlling for nearly all imaginable variables, the researchers, a six-member team led by psychologist Charles Holahan of the University of Texas, found that over a 20-year period, mortality rates were highest for those who were not current drinkers, regardless of whether they used to be alcoholics, second highest for heavy drinkers and lowest for moderate drinkers.

The sample of those who were studied included individuals between ages 55 and 65 who had had any kind of outpatient care in the previous three years. The 1,824 participants were followed for 20 years. One drawback of the sample: a disproportionate number, 73%, were men. Just over 69% of the abstainers died during the 20 years, 60% of the heavy drinkers died and only 41% of moderate drinkers died.

These are remarkable statistics. Even though heavy drinking is associated with higher risk for cirrhosis, stomach ulcer, and several types of cancer, particularly cancers in the mouth and
esophagus, heavy drinkers are less likely to die than people who don't drink even after controlling for all other relevant factors mentioned before.

How can the result be explained? According to the researchers, one crucial reason is that alcohol lubricates so many social interactions. Social interactions are vital for maintaining mental health, which, in turn, contributes to improvement of physical health. As Dr. Holahan pointed out last year, nondrinkers show greater signs of depression than those who allow themselves to join the party.

The authors of the new paper are careful to note, however, that even if drinking is associated with longer life, it can be dangerous. Heavy drinking can impair your memory severely and it can lead to nonlethal falls and other mishaps that can negatively affect your life, such as cheating on a spouse. There is also the dependency issue: if you become addicted to alcohol, you may spend a long time trying to get off the bottle. However, the authors also said that the new study provides the strongest evidence that moderate drinking is not only fun but good for you. So make mine a double.
APPENDIX B
SAMPLE QUESTIONS FOR THE READING SKILL TEST

A sample memory-based question

How many drinks per day are considered as moderate alcohol consumption?

1. 5 drinks for men and 3 drinks for women
2. 5 drinks for men and 2 drinks for women
3. 3 drinks for men and 2 drinks for women
4. 3 drinks for men and 1 drink for women

A sample inference-based question

Which statement is NOT correct about the results and interpretations of the study introduced in the text?

1. Many abstainers who had been former heavy drinkers died early due to serious health problems associated with drinking
2. Abstainers who had not been alcoholic likely died earlier than currently heavy drinkers
3. It is uncertain that abstainers die earlier than current heavy drinkers if they are women
4. Abstainers tended to live longer than current heavy drinkers if they had healthy dietary habits
LIST OF REFERENCES


Linderholm, T., Cong, X., & Zhao, Q. (2008). Differences in low and high working-memory capacity readers' cognitive and metacognitive processing patterns as a function of reading for different purposes. Reading Psychology, 29(1), 61-85.


BIOGRAPHICAL SKETCH

Heekyung Kwon was born and grew up in a suburban area of Seoul, South Korea. She earned her Bachelor of Arts in education and her Master of Arts in educational psychology from the Seoul National University in South Korea, 2002 and 2005, respectively. In fall 2006 she began her doctoral study in the Educational Psychology program at the University of Florida, in Gainesville, Florida and graduated with her Doctor of Philosophy in educational psychology in 2011.