Metacognition Learning (2008) 3:175–188
DOI 10.1007/s11409-008-9025-1

Metacomprehension effects situated within an anchoring and adjustment framework

Tracy Linderholm · Qin Zhao · David J. Therriault · Kristi Cordell-McNulty

Received: 3 June 2008 / Accepted: 3 June 2008 /
Published online: 17 June 2008

Abstract Low accuracy levels are often obtained when readers are asked to predict test performance over reading materials. Three investigations further explore the information readers use to make predictions during metacomprehension. Our results show that readers' estimates are influenced by factors such as their initial impression of the reading task, based in part on their perceptions surrounding text genre and test item type. To explain these and other published results, a new framework for investigating metacomprehension using Tversky and Kahneman's (Science, 185:1124–1131, 1974) anchoring and adjustment heuristic as a guide is proposed. We argue that readers anchor comprehension test performance on factors such as self-perceptions of reading ability and/or perceptions of the reading task and then insufficiently adjust their predictions to reflect the demands of the specific reading task at hand such as text difficulty.

Keywords Metacomprehension · Anchoring and adjustment · Performance predictions · Reading comprehension · Metacognition

Metacognition is a general term used to describe awareness of our own learning, memory, and thought processes (see Flavell 1979; Flavell et al. 1993). Arguably, the more we know what factors make our own learning and memory successful the more likely we are to achieve in academic settings and excel in our professions (e.g., Zimmerman 1989). Unfortunately, research indicates that even some of our most successful learners, college

T. Linderholm (✉) · Q. Zhao · D. J. Therriault · K. Cordell-McNulty
Department of Educational Psychology, University of Florida, 1403 Norman Hall, P.O. Box 117047, Gainesville, FL 32611-7047, USA
e-mail: linderholm@coe.ufl.edu

Q. Zhao
e-mail: qin@ufl.edu

D. J. Therriault
e-mail: therriault@coe.ufl.edu

K. Cordell-McNulty
e-mail: kristicm@ufl.edu

students, have misconceptions about how to facilitate their own learning to prepare for testing situations (Kornell and Bjork 2007). Given our dynamic society and its ever-changing demands on our ability to independently learn, it is critical that we understand the components that go into accurately assessing our own learning (Kornell and Bjork 2007).

In this paper, we narrow our discussion of metacognition to an important subcategory: *metacomprehension.* Metacomprehension refers to the monitoring processes readers engage in when assessing their own level of comprehension. Given that, at all levels of education, educators depend on students to independently learn course concepts from texts, metacomprehension is a process that must be well understood. And with on-line learning (e.g., distance education) becoming more and more prevalent, the ability to assess one's own reading comprehension success is becoming even more crucial (see Maki and Maki 2002).

The standard way to measure metacomprehension is to ask readers to read a series of texts and then ask them to make predictions regarding how well they would do on a test over the texts. This data collection method aligns well with what students naturally do when preparing for college exams. This aspect of metacomprehension, the prediction of comprehension test performance (e.g., Maki et al. 2005) is the focus of the studies reported here. The advantage of adopting this focus is that our studies can be easily compared to previous studies using similar methods and we believe that this focus is more closely attuned to naturalistic academic behaviors. The disadvantage is that two potential processes are being investigated simultaneously: comprehension estimation and one's ability to predict test performance. To lessen the effects of this disadvantage, we clarify that our studies are focused on reading comprehension test predictions, which may involve two processes.

Ideally, predictions accurately reflect students' reading comprehension and match actual test performance. As a strategy, accurately evaluating one's own reading comprehension is key to not only enhancing test performance but to enhancing the overall degree of comprehension (Thiede and Anderson 2003; Thiede et al. 2003) and the first step in the ability to self-regulate learning (Zimmerman 1989). One can imagine that remedial strategies are more likely to be engaged by the reader when one has an accurate sense of future test performance. For example, if a student is able to predict that their learning of textbook material will not earn them the desired grade of "A" in a class, then they are more likely to go back and reread the materials to gain greater comprehension of the material. Given the potential for predictive accuracy to actually enhance learning and school performance, the purpose of this paper is to examine the bases for making predictions of reading comprehension test performance.

To explore the factors that influence metacomprehension predictions, we begin by reviewing the factors that influence readers' comprehension test predictions, report on several studies we have conducted on this topic, and, finally, provide a theoretical framework to consolidate these reported findings on metacomprehension processes.

Prior research on the bases of the prediction process

Some researchers have found that predictions for comprehension performance are influenced by *experiences with the current task* such as familiarity with the text topic (e.g., Glenberg and Epstein 1987; Glenberg et al. 1987; Maki and Serra 1992), ease of text processing (e.g., Dunlosky et al. 2006; Rawson and Dunlosky 2002), and ease of immediate text recall (Morris 1990). For example, Maki and Serra (1992) asked participants to rate their familiarity with text topics solely based on text titles and descriptions. The familiarity ratings were correlated with performance predictions made after reading the full texts. This

finding suggests that *topic familiarity* is a basis of predictions for comprehension performance. In addition to topic familiarity, *ease of text processing* also influences performance predictions. For instance, Rawson and Dunlosky (2002) manipulated ease of processing by using texts with varying degrees of coherence (Experiments 1, 2, and 3) and by using texts with deleted-letter words versus intact texts (Experiment 4). Prediction magnitudes increased as text coherence increased and predictions were higher for intact texts than for texts with deleted-letter words. This finding was replicated by Dunlosky et al. (2006) who found that performance predictions were higher for coherent texts than for the less coherent texts. Performance predictions were also correlated with *ease of immediate recall*, that is, how easy it is to recall text information immediately after reading (Morris 1990). In all, readers may predict comprehension performance based on a variety of experiences with the current task such as topic familiarity, ease of processing, and ease of immediate recall.

Recently, there is also evidence suggesting that performance predictions are largely based on *self-perceptions of ability* (e.g., Moore et al. 2005; Zhao et al. 2006). Specifically, Moore et al. (2005), using a path model analysis, examined the relations among predictions, actual performance, and postdictions within and across three reading trials. Predictions and postdictions are performance estimates before and after a comprehension test, respectively. They found that, despite manipulating text difficulty across trials and giving participants nine learning trials first, performance estimates were fairly stable across the three final trials, that is, trials 10, 11, and 12. For instance, the prediction at trial 10 significantly influenced the predictions at both trial 11 (standardized path coefficient=0.65, $p<0.05$) and at trial 12 (standardized path coefficient=0.42, $p<0.05$). Moore et al. (2005) explained that adult readers such as the undergraduate student participants in this study had gained much experience and skill in reading comprehension tasks, so they might have based performance estimates largely on self-perceptions of ability formed from past experiences. That is, a heavy reliance on self-perceptions of ability as a basis of performance estimates may explain the relatively stable estimates across reading trials.

Another recent study has provided direct evidence that self-perception of ability is a major basis of performance predictions besides experiences with the current task (Zhao et al. 2006). In Zhao et al. (2006), participants read several texts and after finishing reading each text, they predicted their own comprehension performance and then took a comprehension test. Finally, a self-report questionnaire was used that required participants to describe how they made performance predictions. Participants' responses were first categorized and then frequency counts for each category were calculated. This descriptive analysis confirmed that experiences with the current task such as topic familiarity, topic interest, ease of processing, and ease of immediate recall influence predictions. The analysis also revealed another type of prediction basis: self-perception of ability in terms of past comprehension performance. A sample response in this category was, "I used prior knowledge of how well I have done on tests for comprehension in the past where I have read the information directly before the test, as well as what I could understand and recall to myself from the texts right before taking the experiment test." So, previous research shows that experiences with the current task and self-perception of ability serve as bases of comprehension test performance predictions.

Three further investigations of the bases of predictions

In the following section, we present three investigations we have conducted to further examine the factors that influence readers' comprehension test predictions. We primarily

focused our attention on readers' perceptions of the reading task and on the influence of reading skill, as measured by a working memory task. If more is understood about the bases of predictions, then future studies can examine more effectively the factors that increase metacomprehension accuracy.

Study 1: Performance expectations according to text genre and test item type

In the first investigation, we examine how text genre and test item type may affect a person's expectations of comprehension performance out of the context of actual written materials. This study demonstrates the potential variability of performance expectations across different types of materials and test items. We asked research participants to consider how well they expected to do on test items that cover a variety of text genres (e.g., narrative versus expository texts) and test item types (e.g., multiple-choice versus essay test questions). We expected that study participants would be more inclined to rate expository text genres as more difficult because they are often structured in unfamiliar ways and readers are less able to draw on background knowledge to process them in contrast to narrative texts (Kintsch 1998; Wolfe 2005). As for test item types, it was assumed that essay questions would require a deeper mental representation of text information than would multiple choice questions and other items that involved an element of guessing and would, thus, be rated as more difficult (Scouller 1998).

To reiterate, we assumed in this study that since the participants were making predictive estimates completely out of the context of actual written materials that their pre-conceived notions about their abilities to perform under these conditions would serve as a basis for predictions. In other words, the one piece of information that readers focused onto make predictions was the text genre or test item type.

Method Sixty-three college-aged participants from a large southeastern university were recruited to take part in this study as part of a course requirement. Participants filled out a survey in a group format. The survey questions asked participants to rate their expected comprehension performance on a variety of text genres and on a variety of test item types. Specifically, participants rated how well they would do in reading comprehension of narrative texts (e.g., novels), expository texts (e.g., textbooks), and some specific types of expository texts including information reports (e.g., newspaper articles), argument texts (e.g., editorial pieces), and explanatory texts (e.g., instruction manuals). Participants then rated how well they would do on various types of comprehension test items including multiple-choice, true/false, matching, fill-in-the-blank, and essay. All ratings of performance were based on a percentage scale of 5% to 100%, in increments of 5%.

Results The results showed a significant difference in rating levels among the text genres, $F(4, 62)=3.38$, $p=0.01$. Rating levels were in terms of expected percentage of correct answers and ranged from a low of 10% to a high of 100%. An alpha level of 0.0125 was used to determine significance after a Bonferroni correction for four follow-up analyses was applied. Analyses showed the primary difference among pairs of genres was the difference in rating levels between the narrative genre ($M=84.80\%$, SE=1.12%) and the expository genre ($M=79.21\%$, SE=1.78%), $t(62)=3.70$, $p=.001$. Comparisons were also made in terms of the rating levels between narrative text ($M=84.80\%$, SE=1.12%) and each specific type of expository text, including information reports ($M=81.39\%$, SE=1.73%), $t(62)=2.24$, $p=0.028$, argument texts ($M=81.11\%$, SE=1.77%), $t(62)=2.41$, $p=0.019$, and explanatory texts ($M=80.87\%$, SE=2.09%), $t(62)=2.02$, $p=0.047$. However, using the

more conservative alpha level of 0.0125 as a criterion for significance would yield only a significant difference between narrative and expository texts in general. Thus, overall, participants expected their reading comprehension performance to be better on narrative texts than on expository texts.

With regard to test item types, there also was a significant difference in rating levels among them, F (4, 62)=5.57, $p<0.001$. Rating levels in terms of percentage correct with regard to test types ranged from a low of 30% to a high of 100%. Follow-up t-tests, using a significance criterion of.01 for five comparisons based on the Bonferroni correction procedure, showed that participants expected to perform better on matching items (M=90.00%, SE=0.93%) than on the other test items, including multiple-choice items (M=86.35%, SE=1.20%), t (62)= 3.58, p=0.001, essay items (M=84.92%, SE=1.91%), t (62)=2.59, p=0.01, fill-in-the-blank items (M=83.57%, SE=1.60%), t (62)=3.89, $p<0.001$, and true/false items (M=82.86%, SE=1.33%), t (62)=5.50, $p<0.001$. So, matching tests are perceived to be the easiest among various types of comprehension tests. In addition, participants expected their comprehension performance to be better on multiple-choice items (M=86.35%, SE=1.20%) than on true/false items (M=82.86%, SE=1.33%), t (62)=3.32, p=0.002.

Discussion The results of this study showed that students have preexisting expectations of how well they would comprehend different text genres and how well they would perform on different types of comprehension test items. Thus, prior to seeing an actual text or an exam over the text, participants' predictions of comprehension test performance were already biased depending on the text genre or type of test items. And given that these predictive estimates were made out of context of real text materials or test questions, it seems likely that the participants in this study used past experiences of their ability to read these text genres or take these test item types to make predictions. Although some of the findings regarding attitudes towards test item types were counterintuitive, for example, students were equally confident answering multiple-choice test items that capitalize on chance/guessing as they were answering essay test items, the results of this investigation certainly suggest that readers employed estimation points relevant to text genre and test item type to make performance predictions that were based on their past experiences in these contexts.

Study 2: Meta-analysis of performance predictions in the context of an actual reading task

To more specifically determine how individuals' estimates for a particular type of test item, namely, multiple-choice, compared to performance in the context of an actual reading task, a meta-analysis was performed on three previously collected data sets. Multiple-choice test questions were selected because this test item type is one of the most commonly used forms of assessment in studies of metacomprehension and makes comparisons with established findings more compatible (e.g., Weaver and Bryant 1995; Maki 1998; Maki et al. 2005). This meta-analysis allowed a more focused study of reading test predictions in the context of an actual reading task.

Results In three different studies (Linderholm et al. 2008; Linderholm and Zhao 2008; Zhao et al. 2006), college-aged participants were asked to read expository texts and estimate how many multiple-choice test questions they could accurately respond to (out of 10 questions) based on their perceived level of comprehension. A total of 297 undergraduate participants were included in the collapsed sample. The results showed that

participants predicted that they would, on average, answer 6.71 (SE=0.08) questions correctly. In addition, 69.40% of the entire sample estimated that they would answer between 6 and 8 questions accurately. Thus, in terms of estimating multiple-choice test performance over moderately difficult expository texts on various science topics, college students seem to base their answers around an average of approximately 70% correct. Related to this, the average test performance over the same texts across this large sample was 7.32, or 73%, items answered correctly out of 10 (SE=0.12), suggesting that readers in this sample slightly underestimated performance but were fairly accurate.

However, when analyzing the data according to how individuals scored on a working-memory capacity test, a test of individual-differences that is related to reading comprehension skill (e.g., Daneman and Carpenter 1980; Kaakinen et al. 2003), a different pattern emerges. For low working-memory capacity readers, predictions averaged 6.73 out of 10 (SE=0.13) but test performance averaged 6.88 (SE=0.15). For high working-memory capacity readers, predictions averaged 6.76 (SE=0.14) but test performance averaged 7.80 (SE=0.18). So, low working-memory capacity readers were fairly accurate in their test performance predictions whereas high working-memory capacity readers were under-confident and, therefore, less accurate.

Discussion Comparing results from the first and second investigations, we conclude that readers' final estimates depend on the degree of ambiguity present in the estimation task. In the first investigation there was more ambiguity in the estimation task than in the second investigation because actual texts were not present for participants to base their estimations on. In this investigation, the majority (70%) of the participants predicted that they would be 60–80% accurate in answering multiple-choice questions after reading actual expository texts. In the first investigation we described, the majority (62%) of the sample thought they would answer multiple-choice test questions accurately 90–95% of the time but made their predictions out of the context of actual reading materials. It is clear that readers' estimation points are influenced when in context, which is not surprising. The difference between the estimates in the first investigation versus the second investigation is a clear reflection that readers adjusted their estimates to reflect their comprehension of real text materials.

Regarding individual differences, what is so notable is the similarity of predictions between two groups, low and high working-memory capacity readers, that differ so greatly in a wide variety of language skills important for reading (see Just and Carpenter 1992, for a review). This finding suggests that college students in this sample have a common estimation point that is perhaps biased toward "average" performance, which translates to a grade of "C" or 70% at the college level. In this case, estimations were biased towards a common college experience rather than an individual's self-perception of ability. A second possible explanation is that high working-memory capacity readers have a skewed misperception of their ability, which leads them to underestimate their actual performance to the level of low working-memory capacity readers' estimates. The second explanation corresponds with findings in other domains that more-skilled individuals tend to underestimate performance perhaps because they have an unrealistic view of others' (comparable) performance levels (see Dunning et al. 2003). These results highlight that how accurately one predicts reading comprehension performance is directly tied to how adequately one adjusts estimations away from perceived averages. And we found that there are individual differences that may be driving the extent to which readers engage in the adjustment process when making performance estimates. What specific factors may have influenced readers' adjustments is explored further in the next investigation.

Study 3: Self-reports of bases of performance predictions

In this final investigation, readers were asked to read the same two expository texts as were used in the meta-analysis and then to judge their level of learning of the texts and predict test performance. Then readers were asked to report how they went about judging their level of learning and predicting test performance. We conducted this study to obtain self-report measures of *how* readers go about making predictions and to determine the relative frequency of factors that may be used in the prediction process.

Method Thirty-four college-aged participants from a large southeastern university took part in this study as part of a course requirement. The instructions explained that the tasks the participants would complete would be similar to learning material for an upcoming test in one of their classes. While reading the two expository texts, participants were asked to keep in mind that they would be asked to judge how well they comprehended the texts. After reading was complete, participants were asked to describe how they went about judging their performance in two separate tasks.

Task one was designed to elicit responses about how participants made their judgments of learning for each text. The instructions for task one were:

> Now that you have read two texts and judged your level of learning of the texts, please describe in the space below how you made your judgments of learning for each text. That is, how did you go about deciding how well you learned from the texts? Please be as specific as possible when describing how you went about making your judgments.

Task two asked participants to assign grades to their anticipated performance on a test of each text. The instructions for task two were:

> Now that you have described how you went about judging your learning of the texts you read, please estimate for each of the texts what grade you believe you might receive on a test over each text (e.g., A, B, C, D, E). Indicate the estimated grade in the space below. Now, how did you go about estimating your grade? Be as specific as possible when describing how you went about estimating your test grades.

Participants' responses from both task one and task two were placed into six categories that were based on a review of the literature in terms of what factors form the bases of predictions (e.g., Glenberg et al. 1987; Moore et al. 2005; Rawson and Dunlosky 2002; Zhao et al. 2006). The categories were: initial exposure, enduring sense of ability, topic knowledge, topic interest, text difficulty, and type/ difficulty of the test. Responses were placed in the *initial exposure* category when the participant mentioned their initial encounter with the task as the basis of their estimates. Some responses came in the form of comparisons between the first text they encountered and the second text. An example is: "Since I feel like I didn't learn as much information from text 2 as I did from text 1, I guessed that I would receive a lower grade of 'C'." The *enduring sense of ability* category was used when the participant mentioned how well they usually do on reading comprehension tests or their usual level of understanding. An example is: "I am very good at reading comprehension as I have been taking standardized tests for 7 years and reading {comprehension} tests for 12 years." The *topic knowledge* category was used when the participant mentioned taking a class on the subject or mentioned some general knowledge about the subject. An example is: "I went about calculating my grade due to which one I know more about and which one corresponds to what I have heard in class

more about. My major is biology, so the decline of songbirds I have learned multiple reasons for…" The *topic interest* category was used when a participant expressed interest in the topic of the text. An example is: "I was not interested in the information about the moon, so I did not comprehend the information. It was not processed in my brain. On the other hand, I did enjoy the songbirds information somewhat, so I would get a B on an exam." Responses were placed in the *text difficulty* category when the participant mentioned that the text was difficult to understand and/or challenging in some way. An example is: "The moon theories reading was more difficult to learn because it felt more theory by theory and had less of tying ideas together by reasoning in the flow of the reading…" Finally, the *type/difficulty of the test* category was used when the subject mentioned that their estimates were based on what type of test they may receive. An example is: "I think that I would get a B on text 1 if it was multiple choice and an A on text 2 if it was multiple choice….However, if the test was fill-in-the-blank or essay, I think I would do at least one or two letter grades worse on each test."

Two raters separately categorized participants' responses into each of the categories that applied. Cohen's kappa was used to determine the level of agreement. For task one, kappa was.78 (SE=0.05) and for task two, kappa was 0.69 (SE=0.06). The kappa statistics for both tasks are typically categorized as substantial agreement between raters. Disagreements were resolved through discussion.

Results Participants' responses to how they went about making predictions of their levels of comprehension were coded separately for task one and task two. For task one, the results showed that the most prevalent category was topic interest, with 53% of participants responding that this factor influenced their comprehension judgments. Initial exposure to the task was the next most prevalent category with 50% of participants responding that this was a factor in their judgments. Text difficulty was mentioned by 35% of participants and topic knowledge was mentioned by 29% of participants. Lastly, 2.9% of participants mentioned type or difficulty of the test and no participants' responses were coded for enduring sense of ability. For task two, initial exposure to the task was the most prevalent category with 41% of participants' responses coded for initial exposure. Topic knowledge and topic interest were the next most prevalent categories consisting of 35% and 32% of participants' responses, respectively. Eighteen percent of participants mentioned text difficulty and 15% mentioned type or difficulty of the test. Finally, 6% of participants were coded for enduring sense of ability.

Analyses using t-tests were then performed on grades participants predicted they would receive if tested on each text. Grades were converted to percentages for analyses. An A was converted to the typically used college standard of 90%, a B was converted to 80%, a C was converted to 70%, and a D was converted to 60%. If participants assigned themselves a "+" or a combination grade (e.g., A/B), then 5% was added to the lowest assigned grade percentage.

The order of the texts, which was counterbalanced, in one case influenced the grade participants gave themselves on the second text. When "Origins of the Moon" was the first text in the packet of materials given to participants, readers estimated 85% performance (B), t (16)=−6.88, $p<0.001$. When the "Songbirds" text was first in the packet, readers estimated 78% (C+) performance on the "Origins of the Moon". The same effect did not hold for the performance predictions on the "Songbirds" text, $t<1$.

Discussion Across tasks, participants consistently indicated that initial exposure to the task influenced how they went about deciding how well they would perform on a

comprehension test. Related to this point, we also found that readers' predictions of comprehension grades were biased, at least in one instance, by which text they initially received. Specifically, readers' performance predictions (e.g., how they graded themselves) were generally lower for both texts when receiving the perceived harder text about songbirds first. Thus, readers appear to be biased by their initial sense of the layout of the task.

Text difficulty, topic knowledge, and topic interest, as reflected by familiarity and structural aspects of the text, were also factors readers used to make estimates. This fits well with other research regarding factors that influence performance predictions (Glenberg et al. 1987; Metcalfe et al. 1993; Rawson and Dunlosky 2002).

Interestingly, self-perception of enduring ability was not mentioned explicitly by participants as a source of predictions as often as we expected. One potential explanation, although speculative, is that participants' use of past performance is an automatic process that would not be indicated in a methodology like self-reports where participants only report conscious, and less automatic, thoughts (see Ericsson and Simon 1993, for a review on verbal protocol methodology). An alternative explanation is that college students are, generally speaking, considered higher than average learners and they may come to see themselves in that way. Thus, their perceived level of ability would be expected to be 'high' and may not influence judgments. Given the existence of data suggesting otherwise (e.g., Moore et al. 2005), we prefer the former explanation.

Between task one and task two some of the factors became more or less important to readers when estimating performance. Interest in the topic at hand was relatively more important for the first task than the second task but text and test difficulty were more important for making estimates for the second task than for the first task. Although this study is exploratory in terms of what readers may use to make estimates we speculate that by the second task, the more practical matters concerning test performance and grades become most prevalent on the minds of readers and factors such as interest therefore become less relevant because they are less crucial for test performance and grades.

A model of predictive estimates based on anchoring and adjustment

Based on our review of the literature and on the investigations described above, we propose that the processes underlying readers' predictions may be thought of in terms of an *anchoring and adjustment heuristic*. But prior to making our argument regarding how our results can be described in this way, we will define the appropriate terms and outline relevant literature.

According to Tversky and Kahneman (1974), when individuals are asked to make decisions under conditions of uncertainty (e.g., How many questions do you think you could answer correctly on a future test?) they do not employ probability algorithms. Rather, individuals make use of simple cognitive rules of thumb; they use existing information as a starting point, that is, an *anchor,* and then *adjust* until a plausible result is obtained. In the decision-making literature this is typically studied by first asking an individual to provide an estimate to a question that involves a quantity or the probability of an event occurring. The question is presented in such a way that biases the estimate because individuals often overly focus on one piece of information, such as initial numbers, as an anchor point. For instance, when solving the math equation "$8\times7\times6\times5\times4\times3\times2\times1$", students typically guess 2,250 is the answer. In contrast when solving the same math problem in reverse order, "$1\times2\times3\times4\times5\times6\times7\times8$", students typically guess 512 is the answer (Tversky and Kahneman 1974). The answer to both questions is obviously the same and is 40,320.

The important point is that the way that the problem is presented (e.g., larger numbers first) biases students' estimates of the mathematical solution. The initial numbers influence the anchor point and the students adjust inadequately to reach their final estimate. We propose that a similar process of anchoring and adjustment occurs when readers attempt to estimate comprehension test performance.

Anchoring and adjustment processes have been used successfully to describe performance predictions involving other language-related tasks (e.g., Scheck et al. 2004; Scheck and Nelson 2005). For example, a recent judgment-of-learning (JOL) study has demonstrated that some combination of anchoring and monitoring affects the magnitude of immediate JOLs (Scheck et al. 2004). In JOL research, researchers study how accurately individuals can judge their learning of word pairs. Scheck et al. (2004) varied the difficulty of the to-be-learned items and then compared the resulting changes in the magnitude of JOLs and those in the level of recall by comparing the slopes of the lines of best fit. The results indicated that an immediate JOL is both influenced by an anchor point and by on-line monitoring of the item difficulty. In other words, in making immediate JOLs, participants used an anchor point and also adjusted away from the anchor based on their monitoring of item difficulty. Although judging learning of word pairs differs from predicting reading comprehension test performance, this study offers some initial empirical support for our proposal that the anchoring and adjustment mechanism underlies the process of accurately predicting performance in a more complex language task such as reading.

Our proposal is that readers' biased anchoring and adjustment heuristics cause their predictions of actual reading comprehension performance to be inaccurate. Individuals likely use several pieces of information as part of an anchoring and adjustment process: attitudes about test type, knowledge of previous performance, topic interest, domain knowledge, or some combination of these sources of information. This conceptual marriage fits well with reading research that reports generally low metacomprehension accuracy rates (Maki 1998). In the decision-making literature, adjustment effects are difficult to obtain as they require effortful processing, and the causes of insufficient adjustment are beginning to be explored (Epley and Gilovich 2005, 2006). In the reading literature, it is yet unexplored how anchoring and adjustment is used by readers to predict their own performance. We hypothesize that readers employ anchor points and that there are several factors that may influence the degree to which adjustment occurs. We propose that, based on our three investigations reported here, readers typically anchor their predictions based on perceived ability and/or their initial exposure to the task, but insufficiently adjust away from that anchor as a function of text difficulty, topic knowledge and/or interest in the topic.

Recently, Zhao and Linderholm (2008) proposed that the anchoring and adjustment heuristic (Tversky and Kahneman 1974) might be used to describe how experiences with the current task and self-perception of ability work together in influencing performance predictions. According to Tversky and Kahneman's (1974) anchoring and adjustment heuristic, when making judgments under uncertainty people start with an anchor value and then adjust away from it to reach a plausible final estimate. The final estimate is biased toward the anchor value because adjustments tend to be insufficient (also see Epley and Gilovich 2001). Zhao and Linderholm (2008) argued that in metacomprehension research, readers usually are uncertain about the content and difficulty of the comprehension test. So, they may make performance predictions by first anchoring on self-perceived ability level and then adjust away from the anchor to account for experiences with the current task. Due to insufficient adjustments, the final prediction values are biased toward the perceived ability level. This explains the reviewed evidence that performance predictions are

influenced by both experiences with the current task (e.g., Morris 1990; Rawson and Dunlosky 2002) and self-perceptions of ability (e.g., Moore et al. 2005; Zhao et al. 2006) but may be affected by the latter to a greater extent (Moore et al. 2005).

In addition to the prior literature, there is evidence from the three studies reported here that support our claims. With regard to anchoring, we have several pieces of evidence to support our proposal that anchors are influenced by perceived ability or initial task exposure. First, readers' estimates have been shown to be inflexible and relatively stable over several trials (Moore et al. 2005). This suggests that anchoring is at work and that readers use some sense of their ability, likely based on past performance, to anchor their estimates. Second, we have evidence from the text genre and test item type survey that the initial sense of the task can bias anchor points. Even prior to seeing actual reading materials, readers have preexisting expectations of comprehension performance on various text genres and test types. Third, in the final investigation we reported, we found that readers explicitly self-reported that their impression of the first text they read influenced later predictions. We also found that which text came first in the random ordering of the text presentation determined participants' own predictions of performance on the second text. Thus, the first encounter that readers had with the texts influenced their anchoring processes and, finally, influenced their predictions for later texts.

We also argue that several factors influence the adjustment process, which is typically thought to be insufficient. First, readers have been shown to base estimates on how well they know the topic described in the text but may actually be overly confident (Glenberg et al. 1987). Second, we found that compared to the out-of-context survey estimates of readers (our first study), readers in the meta-analysis (our second study) showed evidence of adjustment once they were able to evaluate actual texts themselves, which likely reflected their interests in and knowledge of the topics at hand. Third, we also have self-reports from readers (our third study) that topic knowledge and interest in the text's topic influenced their judgments of comprehension on the same texts used in the meta-analysis (our second study). Thus, readers may initially anchor their performance predictions around preexisting information such as enduring sense of ability or initial sense of task such as text genre, and then adjust their estimations based on immediate experiences from reading particular texts perhaps based on topic familiarity.

Other theories certainly exist to explain metacognitive estimates. Two such theories are transfer appropriate monitoring (TAM) and the accessibility hypothesis (see Dunlosky et al. 2005). TAM proposes that accuracy will increase if the processes and information participants access during their estimations match the processes and information participants retrieve at test (e.g., term-specific estimates paired with term-specific measures of comprehension). The accessibility hypothesis purports that readers make estimates based on the quantity of information immediately available in working memory. This hypothesis emphasizes that making predictions is an inferential process and will only be as accurate as what text information is immediately accessible to participants. Whereas TAM has had some success in explaining inaccurate estimates of performance, it cannot completely explain the information readers use to make more top-down, global estimates of performance (see Dunlosky et al. 2005). And the accessibility hypothesis is similar to the notion of "ease of recall" as described previously (e.g., Morris 1990) and has been demonstrated to not be an entirely accurate cue for metacomprehension. We also suggest that the accessibility hypothesis only represents the bottom-up processes that readers use to make estimates, which is similar to our criticism of TAM. Our proposed framework utilizing anchoring and adjustment allows for both top-down (e.g., self-perceptions of ability) and bottom-up processes (e.g., ease of recall) to account for how readers make

predictions so it is a more complete model than the other two theories (see Dunning et al. 2003). In addition, our model is more in line with current theories of the psychology of reading that describe reading as an interactive process involving both top-down and bottom-up processes (e.g., Kintsch 1998; van den Broek et al. 1999). However, it should be noted that we have not directly tested our model against other theories of metacomprehension and, as such, our proposal that reading comprehension test prediction is a process of anchoring and adjustment is perhaps best used to address specific issues in future research.

Future research

Although the investigations we have presented here are not conclusive, they are suggestive of the notion that readers engage in an anchoring and adjustment type of process when predicting comprehension test performance. If how predictions are made is better understood, educators would have a better chance of training readers to make more accurate estimates and more accurate assessments of their comprehension flaws. As such, we have a proposed agenda for future research topics that is embedded within a framework of the anchoring and adjustment heuristic.

First, it is important to determine what is convincing evidence that anchoring and adjustment is taking place in the context of metacomprehension and to attempt to separate the two processes. That is, quantifying how much or whether adjustment away from a particular anchor value is occurring is a crucial endeavor in this area of study. One measure may be to experimentally manipulate anchor points, as has been traditionally done in the decision-making literature, and to then measure the amount of time it takes readers to reach their final estimates. In this example, the amount of time to make the final decision would be indicative of how much effort went in to the adjustment process.

Second, along the lines of the first research topic, it would be important to know how the factors described previously (e.g., prior topic knowledge, interest) influence the degree to which readers adjust away from or toward either an experimenter-generated or a self-generated anchor point. Using Epley and Gilovich's work (2001, 2004, 2005) as a guide, we would recommend adopting the distinction between types of anchors: self-generated (e.g., students based their judgment upon some enduring feeling of ability or how much they like a text topic) and experimenter-provided (e.g., the suggestion of how well most readers typically do on a test prior to the individual making estimates). Although, according to Epley and Gilovich (2004), we might only expect adjustment to occur when anchors are self-generated. Nonetheless, we would recommend researchers first use a more controlled study of anchors by producing an anchor point themselves to precisely measure the influence of the aforementioned factors on the adjustment process.

Third, it has been argued that how sufficiently individuals adjust way from the anchor, and are hence less biased in estimations, depends on the degree of cognitive effort put forth (e.g., Epley and Gilovich 2005). Thus, if there are constraints on the degree of effort one can put in to making accurate predictions, perhaps due to limitations in cognitive resources, one would anticipate that estimates would be closer to anchor points. For example, readers with low working-memory capacities may be less likely to adjust away from the anchor than readers with high working-memory capacities, which would be interesting to confirm given that working-memory capacity is related to reading skill (e.g., Daneman and Carpenter 1980; Just and Carpenter 1992; Linderholm 2002). Manipulations of cognitive load or strains on cognitive resources would test this hypothesis and explain why readers do not adjust sufficiently.

To conclude, we propose that the study of readers' predictions of actual comprehension test performance should be framed as a process of anchoring and adjustment (Tversky and Kahneman 1974). By carefully examining which factors serve as anchors and, perhaps more importantly, which factors inspire readers to adjust away from the biased anchor point will lead to a more complete understanding of how readers may accurately predict performance.

References

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466. doi:10.1016/S0022-5371(80)90312–6.

Dunlosky, J., Baker, J. M. C., Rawson, K. A., & Hertzog, C. (2006). Does aging influence people's metacomprehension? Effects of processing ease on judgments of text learning. *Psychology and Aging*, *21*, 390–400. doi:10.1037/0882-7974.21.2.390.

Dunlosky, J., Rawson, K. A., & Middleton, E. L. (2005). What constrains the accuracy of metacomprehension judgments? Testing the transfer-appropriate-monitoring and accessibility hypotheses. *Journal of Memory and Language*, *52*, 551–565. doi:10.1016/j.jml.2005.01.011.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science*, *12*, 83–87. doi:10.1111/1467-8721.01235.

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science*, *12*, 391–396. doi:10.1111/1467-9280.00372.

Epley, N., & Gilovich, T. (2004). Are adjustments insufficient? *Personality and Social Psychology Bulletin*, *30*, 447–460. doi:10.1177/0146167203261889.

Epley, N., & Gilovich, T. (2005). When effortful thinking influences judgmental anchoring: Differential effects of forewarning and incentives on self-generated and externally provided anchors. *Journal of Behavioral Decision Making*, *18*, 199–212. doi:10.1002/bdm.495.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, *17*, 311–318. doi:10.1111/j.1467-9280.2006.01704.x.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (rev. ed.). Cambridge, MA: MIT Press.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new era of cognitive developmental inquiry. *The American Psychologist*, *34*, 906–911. doi:10.1037/0003-066X.34.10.906.

Flavell, J. H., Miller, P. H., & Miller, S. A. (1993). *Cognitive Development* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, *15*, 84–93.

Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology. General*, *116*, 119–136. doi:10.1037/0096-3445.116.2.119.

Just, M. A., & Carpenter, P. A. (1992). A capacity hypothesis of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149. doi:10.1037/0033-295X.99.1.122.

Kaakinen, J. K., Hyona, J., & Keenan, J. M. (2003). How prior knowledge, WMC, and relevance of information affect eye fixations in expository text. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *29*, 447–457. doi:10.1037/0278-7393.29.3.447.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

Kornell, N., & Bjork, R. A. (2007). The promise and perils of self-regulated learning. *Psychonomic Bulletin & Review*, *14*, 219–224.

Linderholm, T. (2002). Predictive inference generation as a function of working memory capacity and casual text constraints. *Discourse Processes*, *34*(3), 259–280.

Linderholm, T., & Zhao, Q. (2008). The impact of strategy instruction and timing of estimates on low and high working-memory capacity readers' absolute monitoring accuracy. *Learning and Individual Differences*, *18*(2), 135–143. doi:10.1016/j.lindif.2005.09.002.

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*, 61–85. doi:10.1080/02702710701568587.

Maki, R. H. (1998). Test prediction over text materials. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 117–144). Hillsdale, NJ: Erlbaum.

Maki, R. H., & Serra, M. (1992). The basis of test predictions for text material. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *18*, 116–126. doi:10.1037/0278-7393.18.1.116.

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

Maki, W. S., & Maki, R. H. (2002). Multimedia comprehension skill predicts differential outcomes of web-based lecture courses. *Journal of Experimental Psychology. Applied*, *8*, 85–98. doi:10.1037/1076-898X.8.2.85.

Metcalfe, J., Schwartz, B. L., & Joaquim, S. G. (1993). The cue-familiarity heuristic in metacognition. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *19*, 851–864. doi:10.1037/0278-7393.19.4.851.

Moore, D., Lin-Agler, L., & Zabrucky, K. M. (2005). A source of metacomprehension inaccuracy. *Reading Psychology*, *26*, 251–265. doi:10.1080/02702710590962578.

Morris, C. C. (1990). Retrieval processes underlying confidence in comprehension judgments. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *16*, 223–232. doi:10.1037/0278-7393.16.2.223.

Rawson, K. A., & Dunlosky, J. (2002). Are performance predictions for text based on ease of processing? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *28*, 69–80. doi:10.1037/0278-7393.28.1.69.

Scheck, P., Meeter, M., & Nelson, T. O. (2004). Anchoring effects in the absolute accuracy of immediate versus delayed judgments of learning. *Journal of Memory and Language*, *51*, 71–79. doi:10.1016/j.jml.2004.03.004.

Scheck, P., & Nelson, T. O. (2005). Lack of pervasiveness of the underconfidence-with-practice effect: Boundary conditions and an explanation via anchoring. *Journal of Experimental Psychology*, *134*, 124–128.

Scouller, K. (1998). The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay. *Higher Education*, *35*, 453–472. doi:10.1023/A:1003196224280.

Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, *28*, 129–160. doi:10.1016/S0361-476X(02)00011-5.

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology*, *95*, 65–73. doi:10.1037/0022–0663.95.1.66.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*, 1124–1131. doi:10.1126/science.185.4157.1124.

van den Broek, P. W., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the on-line construction of a memory representation. In H. van Oostendorp, & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Erlbaum.

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

Wolfe, M. B. W. (2005). Memory for narrative and expository text: Independent influences semantic associations and text organization. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*, 359–364. doi:10.1037/0278-7393.31.2.359.

Zhao, Q., & Linderholm, T. (2008). Adult metacomprehension: Judgment processes and accuracy constraints. *Educational Psychology Review*, *20*, 191–206. Retrieved from http://www.springerlink.com/content/104855

Zhao, Q., Linderholm, T., & Therriault, D. (2006). Absolute metacomprehension accuracy: The effects of cue-utilization instruction and working-memory capacity. *Poster session presented at the 2006 American Psychological Association Convention*, New Orleans, LA, August

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, *81*, 329–339. doi:10.1037/0022-0663.81.3.329.