What is the Basis for Self-Assessment of Comprehension when Reading Mathematical Expository Texts?

Magnus Österholm

Umeå University, Sweden

Abstract

The purpose of this study was to characterize students' self-assessments when reading mathematical texts, in particular regarding what students use as a basis for evaluations of their own reading comprehension. A total of 91 students read two mathematical texts, and for each text they performed a self-assessment of their comprehension and completed a test of reading comprehension. Students' self-assessments were to a less degree based on their comprehension of the specific text read, but more based on prior experiences. However, the study also produced different results for different types of texts and when focusing on different aspects of reading comprehension.

Keywords: anchoring; adjustment; metacomprehension; reading comprehension; self-assessment

The purpose of this study was to examine an aspect of metacognition among students when trying to learn something new by reading, specifically when evaluating their own comprehension. In general, aspects of metacognition can be seen as crucial if an aim is to have independent learners that have an awareness of, and take responsibility for, their own learning process. Furthermore, a focus on metacognitive aspects in the teaching and learning can also be directly beneficial for students' learning (e.g., see Dignath & Büttner, 2008). Some aspects of metacognition have been studied through different 'self'-notions, such as self-concept (Guay, Marsh, & Boivin, 2003), self-efficacy (Lent, Lopez, Brown, & Gore, 1996), self-assessment (Brookhart, Andolina, Zuza, & Furman, 2004), self-evaluation (Ross, Hogaboam-Gray, & Rolheiser, 2002), and self-regulation (De Corte, Verschaffel, & Op't Eynde, 2000). Sometimes the same notion can have somewhat different meaning in different studies and different notions sometimes seem to have very similar meaning. However, common for all these notions is that they focus on situations where students in some way reflect on themselves as learners. In this article, focus is on *self-assessment*, a notion here used to refer to *students' evaluations of some aspect of their own knowledge*. More specifically, the situation examined in this study is when students read an expository mathematical text. Within this situation, focus is on characterizing students' self-assessments, concerning what students seem to use as a basis for evaluations of their own reading comprehension.

Students' Self-Assessments of Reading Comprehension

Many studies focusing on self-assessment in reading let students read a text, evaluate their own comprehension, and finally take a test of reading comprehension, often repeated for several texts. The correlation between evaluation and test result, which is sometimes called the calibration of comprehension, is then taken as a measure of the students' ability to monitor their comprehension (Glenberg & Epstein, 1985; Lin & Zabrucky, 1998). In general,

this type of empirical studies seem to show an inability among students to monitor their reading comprehension: Pressley and Ghatala (1990) noticed this for studies during the 1980's, Lin and Zabrucky (1998) for studies during the 1990's, and Dunlosky and Lipko (2007, p. 228) summarized the research area by noting that: "Over two decades of research examining judgments of text learning – or metacomprehension – has consistently demonstrated that people's judgment accuracy is quite poor." These articles report on many studies showing correlations between self-assessment and comprehension, usually of the magnitude .3, but also mention studies where the correlation has been larger. Some of these studies are discussed below.

An evaluation of comprehension depends on some type of *criterion of comprehension* (or standard of evaluation, see Baker, 1985b), for example if you remember the specific wording of a text or if you grasp how different parts of a text are related to each other. Therefore, an explanation of a low or non-existing calibration of comprehension could be that students are using different types of criteria than the ones "embedded" in the questions in the test of comprehension. Therefore, it can be a mistake to focus on a one-dimensional quantitative measure of calibration of comprehension. However, empirical studies focusing on calibration of comprehension often seem to make explicit to the readers what types of criteria will be used by informing them about the types of questions that will be asked in the test. In addition, some studies have also varied the types of questions used and not found any significant differences regarding the level of calibration (e.g., Glenberg & Epstein, 1987; Lin, Moore, & Zabrucky, 2001). A limitation in such studies is that they usually have not utilized theories of reading comprehension that specify different dimensions or components of comprehension (see Wiley, Griffin, & Thiede, 2005). Furthermore, there are studies that have varied aspects of the experimental situation (such as the types of texts used), which have resulted in significant differences in calibration (Lin & Zabrucky, 1998; Schommer & Surber,

1986). In addition, the study of Maki et al. (1990, p. 609) showed that calibration is better "when subjects do more active processing during reading," where the more active processing was created by deleting letters in a text. Other studies have also shown higher accuracy in self-assessment, for example by letting students re-read, summarize, or create keywords after reading a text but before making the self-assessment (see Dunlosky & Lipko, 2007).

It has also been shown that when there is low calibration, students' evaluations sometimes are "based on self-classification as expert or nonexpert in the domain of the text, rather than on an assessment of the degree to which the text was comprehended," that is, they are using a type of *domain familiarity* as a criterion of comprehension (Glenberg & Epstein, 1987, p. 84). More generally, based on their review of studies about self-assessment, Zhao and Linderholm (2008) described two main types of bases for assessment: experiences with the current task and pre-formed expectations. They suggested that both these bases are used in self-assessments, which is discussed more in the next section.

A Framework of Self-Assessment

As mentioned in the introduction, the notion of self-assessment is here used to refer to students' evaluations of some aspect of their own knowledge. Therefore, self-assessment can be seen as one aspect of metacognition, since metacognition generally is characterized as cognition about cognition (Nelson, 1996) and self-assessment is the evaluation of (i.e., cognition about) your own knowledge (i.e., an aspect of your own cognition). Furthermore, self-assessment is often included as one important part of self-regulation. Several descriptions and models of self-regulation distinguish between two main components: monitoring (or evaluation) on the one hand and control (or regulation) on the other (Baker, 1985b; Nelson, 1996; Winne, 2001). Self-assessment can be seen as the crucial component of self-regulation: "The decision to apply regulation strategies is made when an individual becomes aware of his or her comprehension difficulties. Failure to adequately evaluate comprehension thus may

halt the execution of strategy use, resulting in less promising learning" (Lin & Zabrucky, 1998, p. 388).

In this study, self-assessment was not examined as part of self-regulation but focus was on self-assessment in itself and in particular what is used as a basis for making a self-assessment. Zhao and Linderholm (2008) included two types of bases for self-assessments in their framework about how different types of information are combined when performing a self-assessment. The two bases are pre-formed expectations (e.g., self-perception of ability in a certain domain) and experiences with the current task (e.g., ease of processing a text or ease of text recall). In addition, the process of forming an evaluation of reading comprehension includes two steps. Readers first "anchor their judgments on pre-formed performance expectations and then adjust their judgments based on experiences with current tasks" (Zhao & Linderholm, 2008, p. 197). This framework explains empirical results of different kinds; in particular students' reliance on domain familiarity for self-assessments and the relatively low accuracy of self-assessments, since anchoring is seen as primary in self-assessment. However, Zhao and Linderholm saw the need for research that more explicitly test the relationships between anchoring and adjustment, in particular regarding the primacy of anchoring, which was done in the study presented here. That is, the phenomena of *anchoring* and *adjustment* were focused on in the present study, in order to examine how self-assessment relates to different types of prior experiences, in particular more distant experiences (anchoring) and more recent experiences (adjustment).

When discussing *bases* for self-assessment, it is not necessarily the case that an individual actively and consciously thinks about what type of basis to use, since many metacognitive processes can happen at a more unconscious level (Brown, 1985; Fitzsimons & Bargh, 2004). Instead, self-assessment could perhaps better be characterized as *a feeling of* (not) knowing or understanding, a formulation frequently used by for example Flavell (1987),

where the specific basis for such a feeling does not need to be available for conscious reflection. This phenomenon was also evident in a previous study (Österholm, 2006b) where students without hesitation could evaluate their reading comprehension, but had much more difficulty to describe the reasons for (i.e., the basis of) this evaluation. Due to this potential unconscious character of metacognitive processes, the notion of basis for self-assessment was in this study operationalized as a statistical connection between a measure of self-assessment and another variable that describes some type of prior experience. This operationalization is described in more detail in the method section.

A Framework of Reading Comprehension

As evident from a critical analysis of previous empirical studies about self-assessment in reading, "a number of important factors in text comprehension have been largely ignored in research on monitoring understanding of text" (Wiley et al., 2005, p. 409). One such factor is conceptualizing the notion of comprehension, which was done in the present study through a framework of three different components (or levels) of reading comprehension (Kintsch, 1998). The same framework was also used by Wiley et al. (2005, p. 411), who saw it as "the most prominent" of these types of frameworks.

The three different components of a reader's comprehension of a certain text are referred to as the surface component, the textbase, and the situation model (Kintsch, 1998). The surface component refers to when only the specific words and phrases themselves are remembered, and nothing about the *meaning* of the phrases and statements. The textbase represents the meaning of the text, that is, the semantic structure of the text, and it "consists of those elements and relations that are directly derived from the text itself [...] without adding anything that is not explicitly specified in the text" (Kintsch, 1998, p. 103). The situation model is a construction that integrates the textbase and aspects of the reader's prior knowledge, for example in order to create inferences based on the content of the text.

The Present Study

The purpose of this study was to characterize students' self-assessments when reading expository mathematical texts, in particular regarding what students base their self-assessments on when evaluating their reading comprehension.

Research questions:

(RQ1) How do students' self-assessments relate to different aspects of reading comprehension?

(RQ2) How do students' self-assessments relate to prior experiences, of both more recent types and more distant types?

The first question focuses on *to what degree* and *in what way* students' self-assessments are related to their comprehension of the specific text. The question of "to what degree" is similar to what has been examined in many previous studies about self-assessment in reading, regarding accuracy of assessments (calibration of comprehension). The question of "in what way" is about different aspects of comprehension, which takes into consideration that comprehension is not a one-dimensional construct (see previous section about reading comprehension).

The second question focuses on if and how students' self-assessments are related to aspects that are not specific for the reading of a certain text, but on prior experiences somewhat related to the reading of the text. In particular, a recent experience here refers to the reading of a text and taking a test of reading comprehension for that text, before repeating this procedure for another text. In the present study it was examined if and how self-assessment for the second text was influenced by the experiences with the first text, including a potential occurrence of learning (i.e., increase in the accuracy of self-assessment). A more distant experience here refers to the potential use of domain familiarity as a basis for self-assessment, regarding prior experiences and performances in mathematics.

Method

Participants

A total of 91 students, 16-18 years old, from two upper secondary schools in different Swedish municipalities voluntarily participated in this study. Both schools were public schools without any specific selection of students; one school from a medium size Swedish city (20,000-50,000 inhabitants) and one school from a larger Swedish city (50,000-200,000 inhabitants). All students were from the natural science program at the Swedish upper secondary level, which is a national program that has the purpose to lay the foundations for further studies in higher education. All students had completed all mandatory mathematics courses within the natural science program (courses A-D), and some of the students had also completed the voluntary course E.

Procedure

The data in this study consisted of information gathered from a situation when students read two mathematical texts. Before reading the texts, students wrote down their grades from mathematics courses and they answered a questionnaire about their beliefs. The questionnaire included questions about several different types of beliefs, but focus in this study was on the questions about their own ability and level of knowledge in mathematics.

Each student read two texts and worked according to the following procedure for each text:

- Reading of the text.
- Making a self-assessment of reading comprehension.
- Answering questions about the content of the text (i.e., taking a test of reading comprehension). The students could not use the text when answering these questions.

For each student, it was randomly decided which of the two texts was read first, which is of importance since focus in this study was on potential effects the first reading could have on self-assessment of the second text.

The Texts

The two texts used in the present study were chosen considering aspects of length, content, and type. Both texts were approximately one page long. These texts were not as short as texts often used in previous research on self-assessment in reading, where often only a single paragraph has been used (Wiley et al., 2005). Such short texts reduce the possibility to cover more complex topics and also to readily examine all levels of reading comprehension, since sufficient structural complexity is needed in order to "allow for comprehension tests that are not heavily influenced by surface memory" (p. 411). Both texts used in this study had enough structural complexity since they focused on logical connections between ideas presented in them, and they can therefore be described as *explanatory* types of text (see Wiley et al., 2005).

One of the texts was about the concept of absolute value and the other was about the procedure of partial fraction decomposition. Both these topics were new for the participating group of students and chosen in order to be able to focus on the students' reading comprehension and not their prior knowledge about the chosen topics. In addition, these topics are usually included in introductory courses at Swedish university level. The topics are therefore suitable for the participating students since they had studied all mandatory mathematics courses in the natural science program at the upper secondary level, making them eligible for university studies in mathematics.

Through the different topics, the two texts also focused on different types of knowledge. The text about absolute value focused on aspects of conceptual knowledge while the text about partial fraction decomposition focused on aspects of procedural knowledge.

This distinction between different types of texts is seen as relevant and essential since the notions of conceptual and procedural knowledge are central to mathematics (Hiebert, 1986) and also to aspects of reading comprehension (Mills, Diehl, Birkmire, & Mou, 1995).

Both texts used in this study included mathematical symbols, which has shown to be a crucial property of texts regarding students' reading comprehension of mathematical texts (Österholm, 2006a). The text about absolute value was used in two versions, varying the amount of symbols used. Inequalities were described using symbols in one version of the text and with words in the other version. For example, one version used the expression "is greater than or equal to" instead of the symbol "≥". This variation was introduced as a follow-up to a previous study of students' reading comprehension of mathematical texts (Österholm, 2006a), in which significant differences were found between a text version totally without symbols and a version using symbols. However, since no significant statistical differences appeared for the different measures used in the present study, between the groups of students who read the different versions of the text about absolute value, no distinction was made between students who read the different versions.

Measures

Reading comprehension. In this study, a test of reading comprehension consisted of open questions of two different kinds, corresponding to measuring the textbase component and the situation model component respectively. Textbase questions were constructed so that they could be answered with some explicit information given in the text, but in order not to measure pure memorization of the texts (i.e., not to measure the surface component), the questions were not formulated in ways that resembled specific formulations given in the texts and they also frequently asked for a justification for a given answer. Situation model questions were constructed so that some more information or knowledge was needed than what was presented in the text in order to answer the questions correctly, in particular to be

able to apply the information given in the text in a novel situation. This type of question is sometimes referred to as a problem solving question in reading comprehension research and answers are seen as depending on a well-formed situation model (e.g., see McNamara, Kintsch, Songer, & Kintsch, 1996). See Appendix for examples of questions together with examples of analysis of students' answers, which was done according to the procedure described below.

To obtain a quantitative measure of reading comprehension, points were given to the answers based on how complete and correct the answers were in relation to the content of the text. For each text and each component of comprehension, points from different questions were summarized. For the text about absolute value, each participant could get a total of 0-12 points as a measure of the quality of the textbase and 0-20 points as a measure of the quality of the situation model. For the text about partial fraction decomposition, the corresponding numbers are 0-8 points and 0-11 points respectively. Guidelines for assigning points to different types of answers were used, both general guidelines for all questions in order to assign points in an equal manner across different questions and specific guidelines for each question in order to assign points in an equal manner across different students.

Self-assessment. Three different questions were given to students for their self-assessment after reading a text: (1) on a four-point scale how much of the content they thought they had understood, from almost nothing to almost everything, (2) how many of five questions about the content of the text they thought they would be able to answer correctly, and (3) how many of five tasks related to the content they thought they would be able to solve. These scales were combined to create a more reliable measure of self-assessment. First, 0-3 points were assigned to the different answers to question (1). For questions (2) and (3), 0 points was assigned when choosing zero or one question/task, since this can be seen as corresponding to answering *almost nothing* on question (1). Similarly, 3 points were assigned

when choosing four or five questions/tasks, while the middle responses were assigned 1 and 2 points respectively. The points from all three questions were then summarized, and each student could therefore get a total of 0-9 points as a measure of their self-assessment.

Some have argued against letting students make the type of global self-assessment used in this study, since different grain sizes are then used in the self-assessment and in the test of reading comprehension, and studies have shown much higher accuracy in students' self-assessments when matching the grain sizes by asking students to judge how well they would be able to recall a certain part of the text (Dunlosky & Lipko, 2007). However, when students study a text on their own, it is reasonable that they have to rely on more global assessments, when thinking about if they have understood the text, for example before continued studying and decisions in self-regulation. Therefore, the method utilizing a more global self-assessment is ecologically valid.

Basis for self-assessment. Different types of potential bases for self-assessment were in the present study examined that relate to *anchoring* and *adjustment*. The bases analyzed in this study are schematically depicted in Figure 1, and described in more detail below.

As possible *anchors*, students' results from mathematics courses, through their grades and number of courses taken, together with beliefs about their own ability and knowledge in mathematics (shortly labeled as self-efficacy) were included (arrows B1 and B2 in Figure 1). Grades were included by summarizing their grades from mathematics courses A to C at the upper secondary level. The three grade levels were given 0-2 points respectively (for Pass, Pass with distinction, and Pass with special distinction), and the students could therefore get a total grade value of 0-6. All students had also completed mathematics course D but few had received their grades at the moment of participating in this study, and this course was therefore excluded when summarizing the grades. Some of the students had completed also mathematics course E, and this fact was therefore included through a dichotomous variable:

whether a student had taken this course or not (coded 1 and 0 respectively). Finally, students responded to five statements about their own ability and knowledge in mathematics (self-efficacy), where they answered on a four-point Likert scale to what degree they agreed with each statement. The questions were about the difficulty of learning mathematics, the need to study before a test, the ability to explain mathematics to peers, how easy the teacher's explanations are, and whether the student sees her/him-self as good at mathematics. By assigning 0-3 points for the answers on each statement, each student could get a total of 0-15 points as a measure of their self-efficacy.

As possible bases for *adjustments*, measures relating directly to the text currently read were included (arrows R1 and R2 in Figure 1), through both measures of reading comprehension (textbase and situation model).

In addition, aspects of the experience of reading the first text could be used as basis for self-assessment for the second text. This type of experience is here seen as potentially acting both as a type of *anchor* and as a type of *adjustment*. On the one hand, this experience could be utilized in a more general, and perhaps superficial, manner by primarily basing the self-assessment of the second text directly on experiences of the first text, here measured through the variables of self-assessment and reading comprehension of that text. This type of utilization is seen as a sign of anchoring since aspects of the experience of reading the first text is not utilized in a manner specific for the experience of reading the second text (arrows PS and PR in Figure 1). On the other hand, the experience of reading the first text could be utilized in order to better understand what type of comprehension is focused on during this experimental situation, so that some kind of learning, or at least adjustment, occurs regarding the self-assessment. In particular, it is of interest when self-assessment for the second text is based on the specific experience of reading the second text (i.e., the measures of reading comprehension of the second text; arrow R2 in Figure 1) more strongly than the self-

assessment of the first text is based on the measures of reading comprehension for the first text (arrow R1 in Figure 1). This is seen as a sign of adjustment since aspects of the experience of reading the first text is utilized in a manner specific for the experience of reading the second text.

Statistical Analyses

In order to analyze the bases for self-assessment, regression analyses were performed with self-assessment as dependent variable and with several different types of independent variables. In particular, three groups of variables were used as predicting variables (see Figure 1): (1) reading comprehension of the current text (textbase and situation model), (2) background variables (grades, courses taken, and self-efficacy), and, only available for the second text being read, (3) reading comprehension and self-assessment of the previous text. A series of hierarchical regression analyses, with these three groups of variables entered into the regression model in sequence, were performed to examine the relationships between these three types of variables regarding their connection to self-assessment. Furthermore, in all regression models, the relevancies of singular variables within the three main groups of variables were also examined, through the significances of regression coefficients. In particular, it was examined which, if any, of the two components of reading comprehension was statistically related to students' self-assessment.

Results

Students who did not complete all parts of the data collection procedure were not included in the analyses. Therefore, the analyses were based on data from a total of 79 students, of which 38 students first read the text about absolute value and then read the text about partial fraction decomposition, while 41 students read the texts in the opposite order.

Descriptive and Correlation Statistics

The means, standard deviations, and Cronbach's alphas (as measures of internal consistency) are shown in Table 1 and Table 2. All measures except measures of reading comprehension showed good internal consistency. However, it is expected that measures of reading comprehension show lower internal consistency, for two reasons. First, different reading comprehension questions were not intended to measure the exact same construct, but instead to measure different aspects of the reader's comprehension. That is, the comprehension measures could include a property of multidimensionality. Second, sometimes few items were used, in particular for the textbase measure for the text about partial fraction decomposition, where three items were used. The number of items used is an effect of the length of the texts and the type of text used. For a procedural text of one page it was difficult to create several different questions for each type of comprehension measure. And for practical and ethical reasons, the texts were chosen not to be longer than one page, in order for the whole data collection procedure to be reasonably long.

Table 3 and Table 4 show the correlations between all variables included in the study. Each self-assessment variable correlated significantly with all other variables, except on one occasion: For the students who first read the text about partial fraction decomposition, the correlation between self-assessment for this text and the textbase component of reading comprehension for the same text was not significant. However, in order to have symmetry in all comparative analyses, the textbase component was included in all relevant regression analyses.

Because of high correlations between some variables, the risk of multicollinearity was assessed. In all analyses, the variance inflation factors (VIFs) were smaller than five, indicating that multicollinearity was not a problem in the regression analyses.

Predicting Self-Assessment

Three groups of predicting variables were used in regression models, with self-assessment as the dependent variable. The three groups of variables are hereafter shortly referred to using the following numbering and names:

(1) *current text*, which refers to the two reading comprehension variables of textbase and situation model for the current text (arrows R1 and R2 in Figure 1),

(2) *background*, which refers to the three variables of grades, courses taken, and self-efficacy (arrows B1 and B2 in Figure 1), and

(3) *previous text*, which refers to the two reading comprehension variables and the self-assessment variable for the previous text (arrows PR and PS in Figure 1).

Regression models were analyzed for a total of four situations; based on two types of texts (about absolute value and partial fraction decomposition) and whether the text was read first or second. For each type of text, five different regression models were used in the analyses:

- model A1, which included group of variables 1 as independent variables,
- model A2, which included group of variables 2 as independent variables,
- model A3, which included group of variables 3 as independent variables,
- model B, which included groups of variables 1 and 2 as independent variables, and
- model C, which included all three groups of variables as independent variables.

Relations between self-assessment and reading comprehension (RQ1). Around 61% of the variation of self-assessment was explained by reading comprehension variables for the text about absolute value while for the text about partial fraction decomposition 26-42% of the variance was explained by reading comprehension variables (see model A1 in Table 5).

Furthermore, when comparing the variables *textbase* and *situation model* in Model A1 (see β-values in Table 5), in general the situation model variable explained more of the variance of students' self-assessments. However, there was a tendency for the textbase variable to explain more variance when texts were read as second text compared to when read as first text. These results about components of reading comprehension were similar for both types of texts.

Relations between self-assessment and prior experiences (RQ2). For the text about absolute value, there was no clear difference between when the text was read first or second: For both situations, the group of background variables did not add any explanatory power in itself while the variables about the current text did (see Table 6) and the corresponding types of regression models explained similar amount of variation (see R^2-values in Table 5). When examining the singular variables there were some differences between the situations when the text about absolute value was read first or second. In particular, in model B, situation model was significant when the text was read first while self-efficacy was significant when the text was read second (see Table 5). However, these were not radical changes of the magnitudes of the regression coefficients, and several of the values of statistical significance lay close to .05.

For the text about partial fraction decomposition, the results were different than for the text about absolute value: Variables about the current text explained more of the variance of students' self-assessments when the text was read as second text compared to when the text was read as first text (see R^2-values in Table 5), and background variables always added explanatory power (see Table 6). The total explained variance was generally smaller for the text about partial fraction decomposition, at least when the text was read first (see R^2-values in Table 5).

For the regression models when variables for the previous text were added, there was, again, a clear difference between the two types of texts. For the text about absolute value, the variables for the previous text did not add any explanatory power to the regression model while the other variables did, and the opposite was true for the text about partial fraction decomposition (see Table 6). More specifically, it was the previous self-assessment that explained most of the variance of the present self-assessment for the text about partial fraction decomposition (see β-values in Table 5). No singular variable was statistically significant in model C for the text about absolute value, showing that the variables are highly interrelated. However, it was still the reading comprehension variables for the present text that had the highest coefficient values, which were also close to being significant ($p=.07$ for both).

Discussion

An interpretation of the results in relation to anchoring and adjustment when students assess their own reading comprehension reveals the following. For the text about partial fraction decomposition, the anchoring effect was clear, as evident in relation to both background variables and also to the previous text. Adjustment happened primarily based on the experimental situation, since more variance of self-assessment was explained by variables related to the present text when the text was read as second text. For the text about absolute value, there was a weaker anchoring effect, since no genuine explanatory power came from background variables or from the previous text. Adjustment occurred primarily based on the present text and not so much from the experience of first reading another text, since the regression models were similar when the text about absolute value was read as first and second text. These results show clear differences between the two texts, but there are also some similarities: When basing their self-assessment on aspects of the present text, this was primarily done based on the situation model component. In addition, based on experiences

from reading the first text, the students tended to adjust their self-assessment to (also or more) include the textbase component as a basis for self-assessment of the second text.

Based on these interpretations, answers to the research questions can be formulated:

(RQ1) The accuracy of self-assessment, here seen as to what degree self-assessment is related to results on tests of reading comprehension, was different for different types of texts. This relation was more significant for the text about absolute value but weaker for the text about partial fraction decomposition. For both texts, this relation was primarily caused by the relation to the situation model component of reading comprehension. That is, the students tend to base their self-assessment more on the application and use of the text than on the specific content of the text in itself.

(RQ2) More distant experiences (background variables) related to self-assessment differently for different types of texts, where this relation was more significant for the text about partial fraction decomposition but weaker for the text about absolute value. More recent experiences (through the previous text read) related to the self-assessment of the text about partial fraction decomposition in two ways: both more superficially, by virtually repeating the same self-assessment for both texts, and in a more relevant manner, when the self-assessment showed a stronger relation to reading comprehension when read as second text. For the text about absolute value, there was no clear effect of more recent experiences on self-assessment, but for both texts there was a common tendency of basing the self-assessment more on the textbase component when the text was read as second text compared to when read as first text.

The accuracy of self-assessment in previous research has usually been measured quantitatively through the correlation between self-assessment and measures of reading comprehension, where correlations often have been around .3, which corresponds to around 9% explained variance of self-assessment. In the present study, the explained variance of

self-assessment from measures of reading comprehension was much higher, with 40% as the lowest and consistently around 61% for the text about absolute value. However, the quantitative measures in this study cannot be directly compared with measures in prior research, since focus here was on singular texts for groups of students while traditional measures of accuracy are based on correlations for individual students who have read several texts. Therefore, through the method used in this paper, a reliance of background variables could be "hidden" within measures of explained variance by comprehension variables, since there is a correlation between background variables and comprehension. When the explained variance from comprehension variables was examined in relation to background variables, comprehension variables genuinely explained 11-15% of the variance of self-assessment, which is of similar magnitude as results from accuracy measures in previous research.

A total of 70-75% of the variance of students' self-assessment was explained in this study, which gives a good basis for answering the main research question, concerning what students base their self-assessment on when evaluating their reading comprehension of mathematical texts. Since only 11-15% of the variance could genuinely be explained by comprehension variables, more general types of variables (including the specific background variables examined here and also the overlap between different types of variables) seem to be the most important bases for students' self-assessment. This result is congruent with the framework of self-assessment where anchoring is primary and adjustments are secondary. However, more specific types of answers to the main research question that are valid for mathematical texts in general are not easy to give, since the results have shown different results for different types of mathematical texts. For some mathematical texts, students seem to base their self-assessment more directly on their reading comprehension while for other texts they base their self-assessment to a larger extent on their prior experiences in mathematics. Therefore, the present study shows the importance of examining different types

of texts, which was also stressed by Wiley et al. (2005). However, the present study also stresses the importance of distinguishing between texts within one of the categories of texts given by Wiley et al. (2005). Here this has been done by noting that both texts are explanatory but differ regarding what type of knowledge is in focus: conceptual or procedural. However, it is not evident that this distinction is what causes the differences observed here, and more studies are needed, in particular of texts of the same type but covering different topics.

For some of the results and conclusions, when focusing on differences between the two types of texts, one should be aware of the dynamic relationship between the two texts in this study, since both are read and it is examined how reading one of the texts is affected by first reading the other. That is, some specific results about a certain text when read as the second text might not be about properties of this specific text but more about properties of the other text that was read first. However, this is not a problem for the overarching conclusions which are also focused on here: If the distinction between conceptual and procedural texts is the main cause of differences between the texts in this study, the results show that students seem to have more difficulty to assess their own reading comprehension for procedural texts, since they seem to rely more on background variables for this type of text. A reason for this phenomenon could be that conceptual and procedural texts might usually be used by students in different ways for learning. For example, for procedural texts it could be more common to test some steps in the described procedure on a separate piece of paper, which was (implicitly) not possible during the experimental situation in this study. Or it could be that procedural texts are most often used together with some specific task, where the reading of the text is not about comprehending the text (as measured in this study) but on judging to what degree the description in the text seems possible to use when trying to solve a specific task or problem. Thus, there could be an issue of lower ecological validity in this study for

the situation when students read a procedural text. Therefore it would be relevant to use another type of situation for data collection when studying self-assessment of procedural texts in order to examine if students then produce more accurate self-assessments (in relation to the same comprehension measures).

Besides the importance of different types of texts, this study also highlights the importance of the nature of the comprehension test, which was also stressed by Wiley et al. (2005). In the present study, different components of reading comprehension had different relations to students' self-assessments. For example, the accuracy of students' self-assessments is lower if focus is on the textbase component than if focus is on the situation model component. Therefore, studies could produce different types of conclusions if they use different types of comprehension questions.

References

Baker, L. (1985a). Differences in the standards used by college students to evaluate their comprehension of expository prose. *Reading Research Quarterly, 20*(3), 297-313.

Baker, L. (1985b). How do we know when we don't understand? Standards for evaluating text comprehension. In D. L. Forrest-Pressley, G. E. MacKinnon & T. G. Waller (Eds.), *Metacognition, cognition, and human performance* (Vol. 1, pp. 155-205). Orlando, FL: Academic Press.

Brookhart, S. M., Andolina, M., Zuza, M., & Furman, R. (2004). Minute math: An action research study of student self-assessment. *Educational Studies in Mathematics, 57*, 213-227.

Brown, A. L. (1985). Metacognition: the development of selective attention strategies for learning from texts. In H. Singer & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (3 ed., pp. 501-526). Newark, Del.: International Reading Association.

De Corte, E., Verschaffel, L., & Op't Eynde, P. (2000). Self-regulation: A characteristic and a goal of mathematics education. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 687-726). London: Academic Press.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning, 3*, 231-264.

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychology Science, 16*(4), 228-232.

Fitzsimons, G. M., & Bargh, J. A. (2004). Automatic self-regulation. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 151-170). New York: Guilford Press.

Flavell, J. H. (1987). Speculations about the nature and development of metacognition. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 21-29). Hillsdale, N.J.: L. Erlbaum Associates.

Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*(4), 702-718.

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition, 15*(1), 84-93.

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology, 95*(1), 124-136.

Hiebert, J. (1986). *Conceptual and procedural knowledge: The case of mathematics*. Hillsdale, N.J.: Erlbaum.

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

Lent, R. W., Lopez, F. G., Brown, S. D., & Gore, P. A. (1996). Latent structure of the sources of mathematics self-efficacy. *Journal of Vocational Behavior, 49*, 292-308.

Lin, L.-M., Moore, D., & Zabrucky, K. M. (2001). An assessment of students' calibration of comprehension and calibration of performance using multiple measures. *Reading Psychology, 22*, 111-128.

Lin, L.-M., & Zabrucky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology, 23*, 345-391.

Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(4), 609-616.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43.

Mills, C. B., Diehl, V. A., Birkmire, D. P., & Mou, L.-C. (1995). Reading procedural texts: Effects of purpose for reading and predictions of reading comprehension models. *Discourse Processes, 20*, 79-107.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51*(2), 102-116.

Otero, J. (1990). *Failures in monitoring text comprehension: An explanation in terms of the construction-integration model* (Technical report No. 90-17). Boulder: Institute of Cognitive Science, University of Colorado. Retrieved August 6, 2008, from http://ics.colorado.edu/techpubs/pdf/90-17.pdf

Pressley, M., & Ghatala, E. S. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist, 25*(1), 19-33.

Ross, J. A., Hogaboam-Gray, A., & Rolheiser, C. (2002). Student self-evaluation in grade 5–6 mathematics effects on problem-solving achievement. *Educational Assessment, 8*(1), 43-59.

Schommer, M., & Surber, J. R. (1986). Comprehension-monitoring failure in skilled adult readers. *Journal of Educational Psychology, 78*(5), 353-357.

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology, 132*(4), 408-428.

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives* (2 ed., pp. 153-189). Mahwah, N.J.: Lawrence Erlbaum Associates Publishers.

Zhao, Q., & Linderholm, T. (2008). Adult metacomprehension: Judgment processes and accuracy constraints. *Educational Psychology Review, 20*, 191-206.

Österholm, M. (2006a). Characterizing reading comprehension of mathematical texts. *Educational Studies in Mathematics, 63*, 325-346.

Österholm, M. (2006b). Metacognition and reading - Criteria for comprehension of mathematics texts. In J. Novotná, H. Moraová, M. Krátká & N. Stehlíková (Eds.), *Proceedings 30th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 289-296). Prague: PME. Retrieved December 14, 2010, from http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-14118

Appendix

The texts and all descriptions were originally in Swedish, and the excerpt and descriptions below have been translated for this article.

Excerpt from the text about absolute value:

From the definition it also follows that

$$|x - y| = x - y \text{ if } x - y \geq 0 \text{ i.e. if } x \geq y$$

$$|x - y| = y - x \text{ if } x - y \leq 0 \text{ i.e. if } x \leq y$$

where *x* and *y* are real numbers. On a number line, this means that $|x - y|$ is the *distance* between the points *x* and *y* independently of how they are positioned in relationship to each other.

Textbase question related to the given excerpt:

If it is true that $|a - b| = b - a$, what relationship must then exist between the numbers a and b? Justify your answer!

Situation model question related to the given excerpt:

Continue the following sentence, and justify the continuation you give:

Since $|x + 7| = |x - (-7)|$ *then* $|x + 7|$ *can be interpreted as the distance…*

General guidelines for giving points to different types of answers:

- 1 point: Answer relates to *part of* some single statement from the text, which allows partly faulty answers, or the answer generalizes in a faulty manner from some single statement from the texts.
- 2 points: Answer relates correctly to some single statement from the text or is based on relationship/reasoning with faulty explanation/description.

- 3 points: Answer is based on relationship/reasoning (e.g., relating to several statements from the text) but where the explanation/description is not completely specified.
- 4 points: Answer is based on fully specified relationship/reasoning.

Specific guidelines for the textbase question given above, based on different types of answers:

- 1 point: *b* is larger than *a*
- 2 points: *b* is larger than or equal to *a*
- 3 points: answer for 1 point together with justification, with reference to the fact that absolute values are always positive or that they refer to distances, or (indirectly) referring to the definition (e.g., by noting that it should be equal to *a*-*b* if *a*>*b*)
- 4 points: answer for 2 points together with justification as described above

Specific guidelines for the situation model question given above, based on different types of answers:

- 1 point: between *x* and 7
- 2 points: ---
- 3 points: between *x* and -7
- 4 points: answer for 3 points together with justification, with reference to the meaning/definition of $|x - y|$

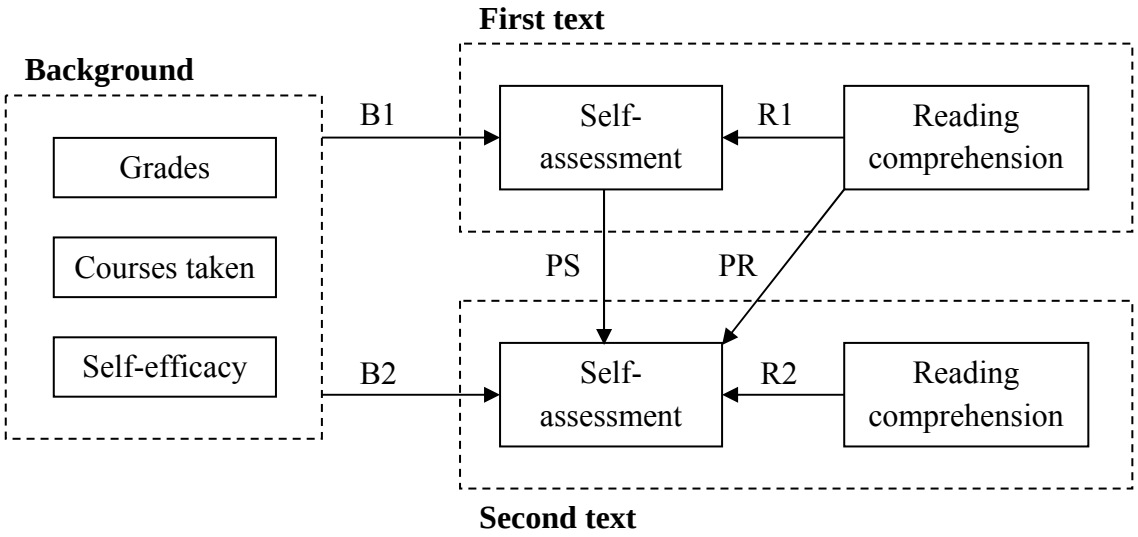

Figure 1. Relationships between variables included in analyses. Arrows point from independent variables towards dependent variables, as included in regression analyses.

Table 1

Descriptive Statistics for All Variables for the 38 Students who First Read the Text About Absolute Value and Thereafter the Text About Partial Fraction Decomposition

Variable	Mean	SD	Cronbach's alpha
1. Grades	7.34	1.43	0.766
2. Courses taken	0.48	0.51	-
3. Self-efficacy	8.74	2.88	0.790
4. Textbase AV	5.16	3.43	0.655
5. Situation model AV	6.98	5.30	0.758
6. Self-assessment AV	5.30	2.89	0.894
7. Textbase PFD	3.10	2.25	0.391
8. Situation model PFD	3.93	3.42	0.697
9. Self-assessment PFD	4.79	3.06	0.905

Note. AV = the text about absolute value; PFD = the text about partial fraction decomposition.

Table 2

Descriptive Statistics for All Variables for the 41 Students who First Read the Text About Partial Fraction Decomposition and Thereafter the Text About Absolute Value

Variable	Mean	SD	Cronbach's alpha
1. Grades	7.02	1.72	0.823
2. Courses taken	0.50	0.51	-
3. Self-efficacy	8.83	2.89	0.733
4. Textbase AV	5.17	3.56	0.618
5. Situation model AV	5.26	4.77	0.754
6. Self-assessment AV	3.78	3.53	0.950
7. Textbase PFD	2.04	2.01	0.397
8. Situation model PFD	3.96	3.63	0.773
9. Self-assessment PFD	6.09	2.81	0.896

Note. AV = the text about absolute value; PFD = the text about partial fraction decomposition.

Table 3

Correlations Between All Variables for the 38 Students who First Read the Text About Absolute Value and Thereafter the Text About Partial Fraction Decomposition

Variable	1	2	3	4	5	6	7	8	9
1. Grades	—								
2. Courses taken	.14	—							
3. Self-efficacy	.66	.34	—						
4. Textbase AV	.59	.42	.46	—					
5. Situation model AV	.53	.55	.58	.71	—				
6. Self-assessment AV	.52	.57	.59	.64	.79	—			
7. Textbase PFD	.10	.27	.26	.26	.19	.32	—		
8. Situation model PFD	.23	.48	.35	.36	.54	.46	.40	—	
9. Self-assessment PFD	.32	.55	.54	.49	.67	.80	.44	.60	—

Note. Correlations of magnitude .32 were significant at $p < .05$. AV = the text about absolute value; PFD = the text about partial fraction decomposition.

Table 4

Correlations Between All Variables for the 41 Students who First Read the Text About Partial Fraction Decomposition and Thereafter the Text About Absolute Value

Variable	1	2	3	4	5	6	7	8	9
1. Grades	—								
2. Courses taken	.26	---							
3. Self-efficacy	.61	.29	---						
4. Textbase AV	.45	.57	.49	---					
5. Situation model AV	.53	.60	.51	.73	---				
6. Self-assessment AV	.44	.59	.58	.72	.73	---			
7. Textbase PFD	.37	.19	.35	.38	.35	.40	---		
8. Situation model PFD	.61	.43	.59	.65	.64	.54	.56	---	
9. Self-assessment PFD	.49	.52	.49	.52	.61	.60	.23	.53	—

Note. Correlations of magnitude .31 were significant at $p < .05$. AV = the text about absolute value; PFD = the text about partial fraction decomposition.

Table 5

Regression Models for Predictions of Self-Assessment of Reading Comprehension

		Text about absolute value				Text about partial fraction decomposition			
		Read first (N=38)		Read second (N=41)		Read first (N=41)		Read second (N=38)	
Model	Variable	R^2	β	R^2	β	R^2	β	R^2	β
A1		.61		.62		.26		.42	
	Textbase current		.15		.39*		–.08		.24
	Situation model current		.66***		.45**		.55**		.51**
A2		.56		.54		.39		.45	
	Grades		.30*		.07		.23		–.01
	Courses taken		.47***		.45***		.36**		.41**
	Self-efficacy		.23		.41**		.23		.41*
A3				.46				.65	
	Textbase previous				.18				–.08
	Situation model previous				.20				.13
	Self-assessment previous				.46**				.75***
B		.67		.67		.40		.57	
	Grades		.11		–.07		.19		–.01
	Courses taken		.22		.17		.31*		.25
	Self-efficacy		.18		.27*		.20		.31
	Textbase current		.09		.28		–.08		.17
	Situation model current		.43*		.33		.17		.31*
C				.70				.75	
	Grades				–.08				–.22
	Courses taken				.15				–.02
	Self-efficacy				.27				.18
	Textbase current				.30				.11
	Situation model current				.31				.25
	Textbase previous				.16				.00
	Situation model previous				–.20				.01
	Self-assessment previous				.15				.66**

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 6

Values of ΔR^2 in Hierarchical Regression Models When Predicting Self-Assessment of Reading Comprehension

		Text about absolute value		Text about partial fraction decomposition	
Models	Group(s) of variables added	Read first (N=38)	Read second (N=41)	Read first (N=41)	Read second (N=38)
A1 → B	Background	.07	.06	.14*	.15*
A2 → B	Current text	.11**	.14**	.01	.12*
B → C	Previous text		.03		.18**
A3 → C	Background & Current text		.25**		.11

* $p < .05$. ** $p < .01$.