How to assess and improve children's reading comprehension?

Muijselaar, M.M.L.

Citation for published version (APA):
CHAPTER 7

GENERAL DISCUSSION
The main goal of this dissertation was to gain further insight into the assessment and improvement of children’s reading comprehension. In what follows, the findings on reading comprehension assessment are reviewed first and discussed in relation to the underlying skills of reading comprehension. Then, the improvement of reading comprehension is described. More specifically, the role of knowledge of reading strategies in reading comprehension is summarized. Last, practical implications and recommendations for future research are given.

**THE ASSESSMENT OF READING COMPREHENSION**

Reading comprehension is typically measured with tests that require children to read a text and to answer questions about it. Reading comprehension measures are often used interchangeably, which implicitly assumes that these tests are measuring the same construct, or in other words, that reading comprehension is a one-dimensional ability. However, several previous studies have shown that reading comprehension tests differ in the cognitive skills that are required (e.g., Andreassen & Bråten, 2010; Keenan, Betjemann, & Olson, 2008). In this dissertation, the demands of several reading comprehension tests were compared in Chapters 2 and 3, and the dimensionality of reading comprehension was examined in Chapter 4.

In the study reported in Chapter 2, two Dutch reading comprehension tests were compared. The results of this study showed that the effects of updating ability and knowledge of reading strategies were similar for the two measures (i.e., the LOVS and the CLIB tests; Leerling- en Onderwijs Volgsysteem, [System for the Longitudinal Assessment of School Achievement]; Cito, 2008; 1992). In contrast to previous studies (e.g., Keenan et al., 2008), no differences between these tests were found in the cognitive skills that are required. One explanation is that the difference between the LOVS and the CLIB tests is too small. The CLIB test consists of cloze items, but although the LOVS test mainly contains question-and-answer items, it also has some cloze items. However, additional analyses showed that when pure cloze and question-and-answer measures were constructed, there were also no differences between the tests in the relations with the cognitive skills. Another explanation for the fact that the findings in Chapter 2 are not in line with previous studies is that the
tests in previous studies did not only differ in question format, but also with respect to text length (e.g., Keenan et al., 2008). In the current study, the passages were relatively long, whereas passages of some tests used in previous studies consisted of one or two sentences only. Thus, differences between reading comprehension tests in previous studies might also be attributed to the fact that reading comprehension tests are used with passages consisting of single sentences only.

In Chapter 3, two American reading comprehension tests, the CBM-Maze and the Gates-MacGinitie reading comprehension tests (Deno, 1985; Espin & Foegen, 1996; MacGinitie, MacGinitie, Maria, & Dreyer, 2000), were compared on their demands. The results of this study showed that the CBM-Maze and the Gates-MacGinitie tests correlated substantially with each other ($r$ ranged from .69 to .79), as well as with word reading skills ($r$ ranged from .47 to .88) and language comprehension abilities ($r$ ranged from .49 to .76). However, when the specific demands of both tests were compared, the CBM-Maze test was found to rely more on word reading skills than the Gates-MacGinitie test. More importantly, the CBM-Maze test requires less language comprehension skills than the Gates-MacGinitie test.

Compared to the relatively homogeneous Dutch tests in Chapter 2, the American tests in Chapter 3 are much more different. First, the CBM-Maze test is timed, whereas the Gates-MacGinitie is untimed. Timed tests might depend more heavily on word reading fluency than untimed tests. Second, the Gates-MacGinitie test consists of question-and-answer items, whereas the CBM-Maze test contains cloze items. These cloze items are constructed by omitting each seventh word from a text and replacing it with three multiple choice alternatives: one correct word and two clearly incorrect words. For example: ‘Was/Ice/This was the second time in three mix/who/days that Holy Angels grads helped one of/not/how their own’. This example shows that the clearly incorrect words are often syntactically incorrect and therefore do not require higher-order comprehension processes. The way the items of the CBM-Maze test are constructed, might thus result in a test that primarily measures lower-order comprehension processes (i.e., sentence comprehension) (e.g., Gellert & Elbro, 2013). That might explain why the reading fluency demand of the CBM-Maze test is larger than the fluency demand of the Gates-MacGinitie test, as well as why the CBM-Maze test hardly requires language comprehension processes.
The study in Chapter 4 went a step further than the studies in Chapters 2 and 3. Chapter 4 is about specific text and question types, whereas the studies in Chapters 2 and 3 concern intact, whole reading comprehension tests. Very early studies revealed that reading comprehension is a one-dimensional construct and, as a consequence, specific text and question types cannot be distinguished (Davis, 1944; Spearritt, 1972; Thorndike, 1973). Previous studies on differences between reading comprehension tests, however, often attributed differences between tests to differences in text and question types (e.g., Keenan et al., 2008). Thereby, these studies implicitly assume that reading comprehension is a multi-dimensional construct. In the study of Chapter 4, multi-trait, multi-method modeling revealed that specific text and question types cannot be distinguished. This study thus confirmed the very early studies on the dimensionality of reading comprehension in showing that reading comprehension seems to be merely a one-dimensional construct. As a result, reading comprehension text and question types do not differ in their cognitive correlates.

To sum up, the most important conclusion of this dissertation with respect to the assessment of reading comprehension is that reading comprehension seems to be mainly a one-dimensional ability (Chapter 4). Specific types of reading comprehension cannot be distinguished. In addition, reading comprehension tests do not differ in the underlying skills that are involved if these tests consist of several texts with questions, and children get enough time to read the texts and answer the questions. Note, however, that there might be differences between the correlates of reading comprehension measures when timed tests or tests that consist of passages of one or two sentences only are compared with standard reading comprehension tests (as in Chapter 3). In addition, it should be noted that most of the results on assessment in this dissertation are focused on Grade 4 children. Since previous studies revealed that the demands of reading comprehension become more complex when children grow older (e.g., Goldman, 2012, Sabatini, O’Reilly, Halderman, & Bruce, 2014), it might be that reading comprehension evolves in a multi-dimensional construct in older children. However, the results of the study in Chapter 3 hardly showed any changing developmental patterns in the word reading and language comprehension skills that are required for reading comprehension in Grades 4, 7, and 9.
The results of the studies in this dissertation imply that reading comprehension tests or specific measures of reading comprehension in previous studies (e.g., Eason, Goldberg, Young, Geist, & Cutting, 2012; Keenan et al., 2008; Kendeou, Papadopoulos, & Spanoudis, 2012) might differ because the reading comprehension tests in those studies differ more strongly (as in Chapter 3) than standard reading comprehension tests as used in the other studies in this dissertation (Chapters 2 and 4). That is, the tests in previous studies and in Chapter 3 are timed or consist of passages of one or two sentences only. It might also be that the differences found in previous studies are accidental differences, because in most of these studies (except for Kendeou et al., 2012), differences in the correlations among specific reading comprehension measures and cognitive abilities were not tested on significance.

**THE UNDERLYING SKILLS OF READING COMPREHENSION**

The studies in this dissertation again showed the role of several underlying lower- and higher-order abilities in reading comprehension. The studies reported in Chapters 2, 3, and 4 showed that word reading skills and reading comprehension are moderately to highly correlated. Additional analyses on the data of Chapter 2 showed that when controlling for vocabulary, reading fluency is an important correlate of reading comprehension in Grade 4. The study reported in Chapter 3 revealed that decoding did not have a unique effect on reading comprehension in Grades 4, 7, and 9. In contrast, reading fluency did contribute to reading comprehension, and this effect remained relatively stable across grades. The results of these chapters are in line with the study of the Language and Reading Research Consortium (2015a), which showed that the role of reading accuracy in reading comprehension will gradually be taken over by reading fluency in children from Grade 1 to Grade 3. Also, Grade 4 has been found to be the critical point for the transition from learning to read to reading to learn (McMaster, Espin, & van den Broek, 2014). Beyond word reading skills, listening comprehension abilities are important for reading comprehension. Vocabulary measures are often used as a proxy for listening comprehension (e.g., de Jong & van der Leij, 2002). The studies reported in Chapters 2, 3, and 4 showed high correlations between vocabulary and reading comprehension, which is also in line with previous studies (e.g., de Jong & van der Leij, 2002; Oakhill & Cain, 2012;
Verhoeven & van Leeuwe, 2008). In addition, the results of the studies in Chapters 2 and 3 showed that vocabulary uniquely predicts reading comprehension.

Beyond these lower-order abilities, several higher-order skills are presumed to be involved in reading comprehension. One higher-order skill that is related to reading comprehension is working memory (Daneman & Merikle, 1996). For the measurement of (working) memory, often simple and complex span tasks are used. Simple span tasks, or storage only tasks, require the involvement of short term memory, whereas complex span tasks, or storage and processing tasks, involve the entire working memory system. The study reported in Chapter 2 revealed that short term memory and reading comprehension are weakly correlated ($r = .22$), and short term memory did not have a specific effect on reading comprehension when reading speed and vocabulary were taken into account. In Chapter 4 it was shown that working memory correlated moderately with reading comprehension ($r = .36$), and an additional analysis revealed that there was a small specific effect of working memory on reading comprehension after controlling for reading speed and vocabulary ($\Delta r^2 = .05$). These correlations were comparable with the results of two meta-analyses on the relation between working memory and reading comprehension (Carretti, Borella, Cornoldi, & de Beni, 2009; Daneman & Merikle, 1996).

In addition to short term memory and working memory measures, more recently also executive functions mechanisms have been related to reading comprehension, such as updating ability (e.g., Carretti et al., 2009). Updating might be considered an important skill in reading comprehension, since the activation of prior knowledge and the integration of this knowledge with information from a text is necessary for the construction of a coherent mental model of the text. In Chapter 2, a small correlation was found between updating ability and reading comprehension ($r = .20 - .22$). In addition, it was found that updating did not contribute to reading comprehension, when reading speed, vocabulary, and verbal short term memory were taken into account. These results do not seem to be in line with previous studies, in which differences were found between poor and good comprehenders on updating ability (Carretti, Cornoldi, de Beni, & Romanò, 2005; Palladino, Cornoldi, de Beni, & Pazzaglia, 2001; Swanson, Howard, & Sáez, 2006). However, even a small correlation between updating and reading comprehension might be sufficient to find differences between poor and good comprehenders in updating ability. Possibly, the
situation model updating process is more complex than the working memory updating process that is measured in Chapter 2. This might explain why updating ability did not have a specific relation with reading comprehension.

**THE IMPROVEMENT OF READING COMPREHENSION**

The studies in this dissertation mainly focused on reading strategies as a process that can be trained to enhance reading comprehension (e.g., National Reading Panel, 2000). Since reading strategies are often used unconsciously, it is difficult to measure the use of reading strategies (e.g., Cromley & Azevedo, 2006). Therefore, the studies in this dissertation focused on knowledge of reading strategies. In Chapter 2, the correlation between knowledge of reading strategies and reading comprehension was found to be moderate. In addition, knowledge of reading strategies had a small specific effect on reading comprehension when reading speed, vocabulary, and memory were taken into account. The study in Chapter 5 examined the developmental relations between knowledge of reading strategies and reading comprehension. The correlations between reading strategies and reading comprehension measured in Grades 4 and 5 were moderate to strong. Developmental analyses revealed that there was a small unique effect of knowledge of reading strategies measured in Grade 4 on reading comprehension measured in Grade 5. In addition, reading comprehension measured in Grade 4 also affected knowledge of reading strategies measured in Grade 5. These results thus showed that there is a reciprocal developmental relation between knowledge of reading strategies and reading comprehension.

Since there were moderate to strong correlations between knowledge of reading strategies and reading comprehension, and knowledge of reading strategies was found to uniquely affect reading comprehension at one time point as well as developmentally, in Chapter 6, the effect of a reading strategy intervention for reading comprehension was investigated. Teachers provided an intervention to their whole class using reciprocal teaching. Previous studies that examined the effect of strategy interventions provided by teachers to whole classes are scarce. It might be argued that beginning comprehenders, in particular, may all profit from reading comprehension interventions, and therefore, a whole class approach in
Grade 4 might be necessary. The strategy intervention in this study affected knowledge of reading strategies. At the posttest and follow-up test, children in the intervention group had more knowledge of reading strategies than children in the control group. The intervention, however, did not have an effect on reading comprehension.

In sum, knowledge of reading strategies is an important contributor to reading comprehension (Chapters 2 and 3), however, the specific effect of knowledge of reading strategies on reading comprehension is small. This might explain why the strategy intervention did not affect reading comprehension. Possibly, for a strategy intervention to be effective, it takes more time and effort.

**PRACTICAL IMPLICATIONS**

As said, the studies in this dissertation showed that reading comprehension measures do not differ in the skills they require, except for reading comprehension tests that differ to a large extent. Furthermore, reading comprehension seems to be a one-dimensional ability. These findings imply that different types of reading comprehension do not exist, at least in fourth graders. However, this is not to say that specific types of reading comprehension could not be trained separately anymore. But when specific types of reading comprehension are trained separately, this will probably affect children’s score on this specific type of reading comprehension, but also on other text and question types that are not the focus of the training.

Another implication of the studies in this dissertation concerns the diagnosis of poor comprehenders. Previous studies found that children that are diagnosed as poor comprehenders with a reading comprehension test that mainly relied on decoding, are only half of the time also diagnosed when a reading comprehension test that merely relied on comprehension processes is used (Keenan et al., 2014). In that study, the tests that relied on decoding consisted of passages with one or two sentences only. The comprehension tests that merely relied on comprehension contained several longer texts with different questions. Thus, for the diagnosis of poor comprehenders, untimed reading comprehension tests consisting of relatively long texts accompanied by different questions should be used.
In addition to the assessment of reading comprehension, this dissertation also focused on further improving children’s reading comprehension, in particular via the knowledge and use of reading strategies. Studies in this dissertation revealed that knowledge of reading strategies had a small but significant contribution to reading comprehension. Therefore, instruction and practice of reading strategies should remain a part of education in reading comprehension. An intervention focused on reading strategies however did not improve the level of reading comprehension. This might be interpreted as that apparently, there is no need to change the current curriculum with respect to the amount of time spend on reading strategies. It could also be that to improve the level of reading comprehension with a strategy training, this training should be more intensive for children and teachers.

**RECOMMENDATIONS FOR FUTURE RESEARCH**

The study reported in Chapter 4 revealed that reading comprehension seems to be mainly a one-dimensional ability in fourth graders. The age of children might, however, influence the dimensionality of reading comprehension. Previous studies showed that reading comprehension processes change as children get older (e.g., Goldman, 2012; Sabatini et al., 2014): lower-order skills become less important, whereas higher-order abilities contribute to a larger extent to reading comprehension. Note, however, that in the study reported in Chapter 3, hardly any changes in the developmental patterns of the demands of reading comprehension were found. In line with previous studies, one might suggest that when children are younger, reading comprehension is a one-dimensional ability, but when children get older, it evolves into a multi-dimensional construct, similar to language ability that also becomes multi-dimensional in children from 4 to 8 years old (Language and Reading Research Consortium, 2015b). Future research should therefore focus on the dimensionality of reading comprehension in children at different ages.

The one-dimensionality of reading comprehension might be explained by how reading comprehension tests are constructed (Shanahan, 2014). For reading comprehension tests to be reliable, the items should be highly correlated. Thus, during the construction of reading comprehension measures, items that do not correlate highly with other items will be removed from the test. This diminishes the chance that reading comprehension measured
with such a test is a multi-dimensional construct. Some reading comprehension tests, however, consist of items that are particularly constructed to measure specific types of reading comprehension (e.g., local and global inference questions, Cain & Oakhill, 2014). Future studies might investigate whether items that are intended to measure different types of reading comprehension, can be distinguished reliably.

The study reported in Chapter 3 revealed that the CBM-Maze test relies more on word reading skills than on language comprehension abilities, suggesting that this test should be revised to be adequate for the formative assessment of reading comprehension. The CBM-Maze test has many advantages over standardized reading comprehension tests, that is, it is fast, easy to administer, and inexpensive (Pierce, McMaster, & Deno, 2010). Future studies should therefore investigate how to revise the CBM-Maze test in order to increase its language comprehension demands. Revisions that could increase such demands are answer options that are less different, syntactically correct, and require higher-order comprehension processes. In addition, gaps in the text could be variably spaced instead of placed with a fixed ratio (i.e., each seventh word; Gellert & Elbro, 2013), such that the gaps fall in places where comprehension of (a part of) the text is crucial to choose the correct alternative.

In addition to reading comprehension assessment, the improvement of reading comprehension was an important theme in this dissertation, in particular via the knowledge and use of reading strategies. For the current studies, a questionnaire about how and when to use reading strategies was used. As a consequence, the relations between reading comprehension and knowledge of reading strategies might not generalize to the use of reading strategies. Therefore, future studies should also measure the use of strategies and examine its relation with reading comprehension.

The study reported in Chapter 5 showed that reading comprehension and knowledge of reading strategies are reciprocally related, that is, knowledge of reading strategies has a unique effect on reading comprehension and vice versa. Previous studies on the developmental relations between reading comprehension and its underlying skills often use reading comprehension as outcome measure (e.g., de Jong & van der Leij, 2002; Oakhill & Cain, 2012). Future studies on the developmental relations between cognitive and linguistic
abilities and reading comprehension should therefore also focus on reciprocal relations (see for example, Quinn, Wagner, Petscher, & Lopez, 2015).

Since higher-order skills were found to be important in reading comprehension, but reading strategies only have a small specific effect on reading comprehension, the question remains how the new generation of reading intervention research should be framed. Several previous studies showed that strategy interventions are effective (e.g., National Reading Panel, 2000), however, in these studies little attention is paid to how reading strategies result in an improvement of reading comprehension. It might be that emphasizing visualization strategies more strongly within strategy interventions would result in a larger improvement of reading comprehension (de Koning & van der Schoot, 2013). Probably, it might even be beneficial to involve multiple sensory modalities during text comprehension, such as suggested by theories of embodied cognition.

Reading strategies might indirectly affect reading comprehension (e.g., Cromley & Azevedo, 2007; McKeown, Beck, & Blake, 2009), or even worse, it is suggested that the use of reading strategies hampers reading comprehension because it takes mental resources, and therefore, less attention can be paid to the construction of a situation model (McKeown et al., 2009). Therefore, future reading comprehension interventions should more directly focus on the construction of a situation model. For example, interventions might have a content or questioning approach (e.g., McKeown et al., 2009; McMaster et al., 2012) in which meaning-based questions about the text are asked. However, further research on the timing of the questions (during or after reading) and the effect of timing on students from different grade-levels is necessary. A study of van den Broek, Tzeng, Risden, Trabasso, & Basche (2001) suggests that for a questioning approach in younger children (4th grade) to be effective, the demands that the texts and questions pose on memory resources should be minimalized. A direct focus on the construction of a situation model might also occur when the intervention aims at knowledge building and inference making (Compton, Miller, Elleman, & Steacy, 2014). Note however that this is not to say that reading strategies should no longer be part of the education in reading comprehension, since it is important to learn how to draw inferences and how to make a summary, but the focus should mainly be on the construction of a situation model.