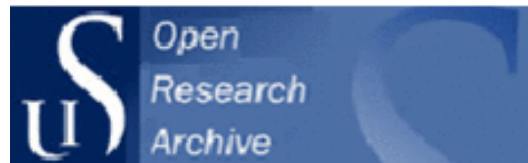




University of  
Stavanger

Solheim, O.J. (2011) The Impact of Reading Self-Efficacy and Task Value on Reading Comprehension Scores in Different Item Formats. *Reading Psychology*, 32(1), pp. 1-27

Link to official URL: [doi:10.1080/02702710903256601](https://doi.org/10.1080/02702710903256601) (Access to content may be restricted)



UiS Brage  
<http://brage.bibsys.no/uis/>

This version is made available in accordance with publisher policies. It is the authors' last version of the article after peer review, usually referred to as postprint. Please cite only the published version using the reference above.



# **THE IMPACT OF READING SELF-EFFICACY AND TASK VALUE ON READING COMPREHENSION SCORES IN DIFFERENT ITEM FORMATS**

**ODDNY JUDITH SOLHEIM**

*It has been hypothesized that students with low self-efficacy will struggle with complex reading tasks in assessment situations. In this study we examined whether perceived reading self-efficacy and reading task value uniquely predicted reading comprehension scores in two different item formats in a sample of fifth-grade students. Results showed that, after controlling for variance associated with word reading ability, listening comprehension, and nonverbal ability through hierarchical multiple regression analysis, reading self-efficacy was a significant positive predictor of reading comprehension scores. For students with low self-efficacy in reading, reading self-efficacy was a significant positive predictor of multiple-choice comprehension scores but not of constructed-response comprehension scores. For students with high self-efficacy in reading, reading self-efficacy did not account for additional variance in either item format. The implication that the multiple-choice format magnifies the impact of self-efficacy in assessments of reading comprehension is discussed.*

## **Introduction**

In order to draw valid inferences from scores on reading comprehension tests, we need to know how and to what extent test characteristics affect individual students' performance. The format used to assess comprehension of written material is one factor determining reading comprehension scores (Fletcher, 2006; Francis, Fletcher, Catts, & Tomblin, 2005). Motivation is another factor influencing students' performance on assessments of reading comprehension (Guthrie & Wigfield, 2005). The aim of the present article is to lay the foundation for informed assumptions about the interaction between item format and motivation in the context

of scores on reading comprehension tests. For this purpose, it views research findings relating to item format and motivation in light of each other. The specific problems under investigation are, first, whether motivation contributes differently to reading comprehension scores depending on item format and, second, whether students with different levels of reading motivation stand to profit from different item formats.

### *Reading Comprehension and Item Format*

Item format has been a recurring subject in the discussion of and research into the assessment of reading comprehension. Because one of the alternatives, multiple choice (MC), is so simple and so economical in terms of testing time and scoring costs, the subject of discussion—particularly in relation to large-scale assessments—has often been whether the use of constructed-response (CR) items, which require students to write their own responses rather than choosing from already formulated alternatives, may be justified despite the extra expense involved. The decisive factor in this discussion has been “value added”: the extent to which the inclusion of CR items increases our capacity to assess reading comprehension in a valid and representative way.

In the late 1980s and early 1990s, during the period characterized by Pearson and Hamm (2005) as a revolution in the field of reading comprehension assessment, the discussion of item format was first and foremost rooted in questions of construct representativeness. In this period, standardized MC tests of reading comprehension were criticized for not reflecting contemporary theories of how readers construct meaning. This criticism drew attention to a number of other aspects: the kind of material presented for students to comprehend; the content of the questions students were asked in order for them to demonstrate comprehension; and limitations inherent in the MC format (Pearson & Hamm; Valencia & Pearson, 1987). As regards item format, it was claimed that CR questions would capture a qualitatively different and deeper form of understanding than MC questions. Answers to CR items were viewed as reading comprehension indicators that were more in line with contemporary theories of reading, meaning that they would contribute to greater consistency between definitions and operationalizations (that is, tests of the construct of

reading comprehension). However, the research on item format has been somewhat inconclusive. There is evidence that students employ different problem-solving strategies when answering MC and CR items, but there is less evidence supporting hypotheses that the CR and MC formats capture different forms of understanding (Alderson, 2000; Campbell, 2005; Pearson & Hamm).

In the 1990s, the CR format attained a major role in the assessment of reading comprehension (Johnson, Jenkins, & Jewell, 2005), and there now seems to be a consensus in the research community about the usefulness of the format. It is generally accepted that measuring the understanding of text by only one method is inadequate (Alderson, 2000) and that the use of multiple measures provides test takers with the opportunity to respond to texts and demonstrate understanding of them in different ways. The CR format also makes it possible to ask questions that could hardly be designed as MC questions (for instance, questions asking students to find textual support for an interpretation). Campbell (2005) stated,

Perhaps it is time, however, to recognize the advantages and disadvantages of both item formats—and begin focus on how each can be better used to maximize the construct representation of reading assessment and to maximize the opportunities for individual test takers to demonstrate their reading abilities. (p. 365)

Campbell's proposed focus on opportunities to demonstrate reading comprehension is the starting point for the research reported in this article. If the CR format makes some individuals better able to demonstrate their potential, then this would be yet another argument in favor of including the CR format in assessments of reading comprehension.

### *Reading Comprehension and Motivation*

We have by now comprehensive evidence for a connection between motivation to read and reading comprehension (Baker & Wigfield, 1999; Guthrie & Wigfield, 2000), and motivation is regarded as a driving force in children's reading development. Further, Guthrie and his colleagues have proved in their research that there is a relationship among motivation to read, amount of

reading, and reading comprehension (Guthrie, Wigfield, Metsala, & Cox, 1999). On the basis of this research, they have developed a model that they call *the engagement model of reading development* (Guthrie & Wigfield, 2000). This model posits that reading comprehension is the result of a large amount of engaged reading. Another example of a theory of reading development in which motivation plays an important part is Alexander's model of domain learning (Alexander, 2005).

A key concept of motivation research is readers' sense that they have the capability to read effectively. This stems from Bandura's (1997) notion of self-efficacy, which he defined as "individuals' confidence in their ability to organize and execute a given course of action to solve a problem or accomplish a task" (p. 3). Based on a number of studies from different domains, Bandura demonstrated how individuals' beliefs about their self-efficacy influence their performance, effort, and persistence as well as their choices of what tasks to perform. A self-efficacious student will participate more readily, work harder, persist longer, and have fewer adverse emotional reactions when encountering difficulties than a student who doubts his or her capabilities. Guthrie et al. (2007) and Zimmerman (2000) have examined the influence exerted by students' confidence in their own reading abilities, finding that students with low reading self-efficacy try to avoid challenging reading activities and tend to withdraw from tasks they perceive as too difficult. In a recent article, Mucherach and Yoder (2008) also found that self-efficacy in relation to reading predicted scores on a standardized reading test in middle school children.

It has been suggested that the connection between motivation and reading comprehension is mediated through strategy use. A series of studies have established a connection between motivation and strategy use (see Pintrich, 2000, for a review). Strategic reading is viewed as a prerequisite for successful reading comprehension (Dermitzaki, Andreou, & Paraskeva, 2008; Pressley, 2002), and in a review of research Carr, Mizelle, and Charak (1998) also argued for a causal connection from motivation to read via reading strategies to reading performance. Strategy use is, however, not the only possible explanation for the observed relationship between motivation and reading. In a study of ninth-grade students, Anmarkrud and Bråten (2009) found that students' beliefs about how important it is to do well on a given

reading task, how useful that task is in relation to current and future goals, and how interesting the individual students find the task to be in and of itself (collectively, reading task value) accounted for additional variance in MC reading comprehension scores after controlling for variance associated with surface strategies (memorization) and deeper strategies (organization, elaboration, and monitoring).

Overall, there is thus evidence that motivation is of importance for reading comprehension development and performance. A related, but different, question is whether the characteristics of reading comprehension tests or test situations affect the motivation of test takers and thereby their performance on assessments of reading comprehension.

#### *Assessment of Reading Comprehension: Item Format and Motivation*

Guthrie and Wigfield (2005) have highlighted several characteristics of tests and test situations that might affect student performance. For the purposes of the present study, the most relevant one is task complexity. They hypothesized that students with low reading self-efficacy will encounter problems when faced with complex reading tasks because “[. . .] if a reading assessment has a high level of complexity, students’ sustained effort, avoidance of distractions, and commitment to completing tasks successfully, are likely to contribute to successful performance” (p. 201). Bandura (1997) has also claimed that skill can easily be overruled by self-doubt, such that skilled persons make poor use of their capabilities under circumstances that undermine their belief in themselves. If certain test characteristics have this effect on some students, and if low scores thus reflect a lack of effort (caused by test circumstances) rather than poor reading comprehension, then this must inform the inferences we make about reading comprehension on the basis of test scores.

It can be assumed on the basis of the research literature that both CR and MC items may add to test complexity. The need to give long written answers is one of the examples given by Guthrie and Wigfield (2005) of the kind of test complexity that may negatively affect the performance of students with low reading self-efficacy. In the present study, the CR questions answered by the participants were short-answer items for which a few words or a

sentence were enough. Even so, writing may still entail a threshold for students in Grade 5. Formulating and writing an answer to a question may be perceived by the participants as more demanding than choosing among alternatives on an MC item, and scores on short-answer CR items can as a consequence be predicted by motivation to a greater extent than scores on MC items.

However, MC items may also be seen to add complexity. Some researchers have argued that, to answer MC items, students must use strategies in addition to the ones activated during reading in nontesting contexts. Farr, Pritchard, and Smitten (1990) studied strategy use during the taking of MC tests in a group of college students, finding little evidence of reading strategies being used while the students were reading the passage for the first time but considerably more use of test-taking strategies when they were answering the questions. Allan (1992) investigated whether the MC format activated different test-taking strategies from the free-response format. His findings were that MC questions activated strategies focusing on the stem (stem is the first part of a MC question appearing before the optional choices) and the alternatives, whereas free-response strategies centred more on the text passage and on students' knowledge of the topic. In addition, Allan found that MC questions activated test-wiseness strategies; for example, the elimination of alternatives or the application of various forms of logical analysis to the structure of the question (see also Alderson, 2000; Rupp, Ferne, & Choi, 2006). A possible conclusion to be drawn from this is that the activation of reading comprehension strategies during the initial reading of a text may influence performance in both the CR and the MC format, whereas the activation of problem-solving strategies may primarily influence performance in the MC format. This means that responses to MC items may be affected by motivational variables to a greater extent than responses to CR items.

*Aim of the Study: Reading Comprehension, Motivation,  
and Item Format*

The aim of the present study is to examine whether motivation predicts reading comprehension scores in two different item formats: multiple choice and short-answer constructed response. The

motivation constructs were added after controlling for word reading ability, listening comprehension, and nonverbal ability. There is a general agreement among researchers that word reading and listening comprehension are important predictors of reading ability (Hoover & Gough, 1990). The reason for including nonverbal ability may not be as obvious. In the present study nonverbal ability was measured by Raven's Standard Progressive Matrices (Raven, 1958; Raven, Court, & Raven 1988). This is a problem-solving task where participants have to reason by analogy, a process that bears resemblance to the way different options are evaluated against each other and the stem while solving an MC question. The measure was included in order to explore whether this kind of problem-solving task predicted MC reading comprehension to a greater extent than CR reading comprehension. The motivation variables are based on expectancy-value theory (Eccles & Wigfield, 2002; Wigfield & Eccles, 2000). The "expectancy" component refers to children's beliefs about how well they will do on upcoming tasks at a domain specific level and is similar to Bandura's (1997) self-efficacy. The expectancy component includes both the individual's ability beliefs in relation to present ability and the individual's own expectations for success in the future (not outcome expectations). The "value" component refers to beliefs about how important it is to do well on a given task, how useful those tasks are in relation to current and future goals, and how intrinsically interesting they are to the individual (Wigfield & Eccles).

The following two research questions were drawn up:

1. Do reading task value and reading self-efficacy make independent contributions to variance in reading comprehension scores over and above the contributions from decoding, listening comprehension, and nonverbal ability in the MC and CR formats, respectively?
2. Do reading task value and/or reading self-efficacy make similar independent contributions, if any, to variance in reading comprehension scores for students at various levels of reading self-efficacy?



## **Method**

### *Participants*

The participants were 217 fifth graders aged 10–11 (51.2% girls and 48.8% boys) from 12 classes at five Norwegian primary schools. Based on scores on national reading tests, the sample was nationally representative in terms of reading comprehension. From an international perspective, the sample was relatively homogeneous as regards socio-economic status (predominantly middle class).

### *Material*

All tests were classroom administered in the pupils' respective classrooms. The reading comprehension tests were administered by the respective class teachers. All other tests were administered by the author and a research assistant.

#### Reading comprehension measures

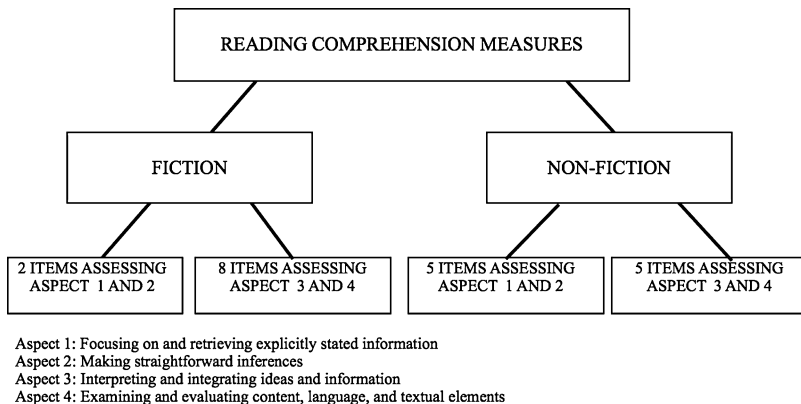
In order to create representative measures of reading comprehension the participants read 11 texts categorized under two main categories: fiction and nonfiction. The texts varied in content, length (from 321 words to 744 words), and complexity. All texts were originally written for children and represented different text types including narrative, short story, report, recipe, instruction, and expository text. The texts were distributed across two booklets, each containing fiction and nonfiction as well as reading comprehension questions. The booklets were administered on two succeeding days, and the participants were given 90 minutes to finish each booklet. Each text was followed by a mixture of MC questions and short-answer CR questions (see Appendixes A and B for sample items). The questions were designed to capture four different aspects of reading comprehension, in line with the four comprehension processes used for item development in the Progress in International Reading Literacy Study (Mullis et al., 2006): (a) focusing on and retrieving explicitly stated information, (b) making straightforward inferences, (c) interpreting and integrating ideas and information, and (d) examining and evaluating content, language, and textual elements

(Mullis, Kennedy, Martin, & Sainsbury, 2006). The participants were allowed to look back at the text passages while answering the questions.

Two measures were constructed: an MC comprehension measure (20 items) and a CR comprehension measure (20 items). Each of the measures included 10 items from fiction texts and 10 items from nonfiction texts. In order to reflect the characteristics of these two main categories of texts, the distribution between different aspects of reading differed somewhat for the two categories. Of the 10 items from nonfiction texts, 5 items in each measure assessed aspect 1 (focusing on and retrieving explicitly stated information) and aspect 2 (making straightforward inferences); the other five items assessed aspect 3 (interpreting and integrating ideas and information) and aspect 4 (examining and evaluating content, language, and textual elements). Of the 10 items from fiction texts, 2 items in each measure assessed aspects 1 and 2 and 8 were designed to assess aspects 3 and 4 (see Figure 1). Scoring criteria for the CR items were written during item development. All items, both MC and CR, were worth 1 point. Reliability (Cronbach’s alpha) was .86 for the MC measure and .85 for the CR measure, indicating a high level of internal consistency.

WORD READING

Participants’ word reading abilities were measured using a standardized word-chain test, *Ordkjedeprøven* (Høien & Tønnesen,



**FIGURE 1** Joint structure of the MC and CR reading comprehension measure.

1998). This is a screening test where participants are instructed to segment letter strings into their constituent words. Four words are combined in each word chain. The words are nouns, verbs, adjectives, adverbs, prepositions, or numerals, and they range in length between two and seven letters. The words in a chain are semantically unrelated (Anglicized example: “treeoverlifese”). There are 90 word chains in the test, and the participants are only allowed to spend 4 minutes on it (in other words, it is impossible to complete all of it). The measure used refers to the number of word chains correctly segmented. The reliability coefficient (Spearman-Brown correction formula based on odd–even correlation) in the standardized sample was .86.

#### Listening comprehension

Listening comprehension was assessed by means of a listening comprehension task that is part of a standardized achievement test in reading for students in Grade 7 (Nasjonalt Læremiddelsenter [NLS], 1995). The task is part of a classroom-administered screening battery whose main purpose is to identify children with special needs in reading at the end of Grade 7. Because of this purpose, several of the tasks have a ceiling effect. In the standardized condition, the text is read out to the children, but they also have the text in front of them and may thus read it themselves. Students then read and answer seven MC questions about the text. In the present study, the task was used on students in Grade 5 and under conditions different from the standardized ones. The participants had no access to the written text, and the text was read to them once. They then read the seven MC questions and answered them in writing. The reliability in the present sample (Cronbach’s alpha) was .72.

#### Non-verbal reasoning

The participants completed all items of Raven’s Standard Progressive Matrices (Raven, 1958; Raven et al., 1988). This is a test designed to measure a person’s ability to form perceptual relations and to reason by analogy, independently of language and formal schooling. In each test item, the student is asked to identify the missing segment required to complete a larger pattern. The booklet used includes five sets (A to E) of 12 items each (e.g., AI

to A12), with the items within a set becoming increasingly difficult. Students completed all five sets and there was no time limit.

#### Inventory of reading motivation

To assess motivation for reading, eight items designed to measure the expectancy component and seven items designed to measure the value component were used. With respect to the latter, two items concerned the importance of reading, four the usefulness of reading, and one the interestingness of reading. Six of the items were adapted from the revised version of the Motivation for Reading Questionnaire (Wigfield & Guthrie, 1997) and seven were adapted from the motivation inventory based on expectancy-value theory presented in Anmarkrud and Bråten (2009). To these were added two newly created items, one pertaining to the value component and one to the expectancy component. All items were read aloud to the students, who rated each item of the reading motivation inventory on a 5-step Likert continuum ranging from *I don't agree* (1) to *I agree* (5). Principal component analysis with orthogonal rotation yielded two components with high loadings and low overlap including all 15 items. The first of these components focused on participants' beliefs about their own reading ability and covered the 8 items originally designed to assess the expectancy component. The second component concerned the value of reading and included the 7 items designed to assess the importance, usefulness, and interestingness of reading. On the basis of this factor analysis, two scales were constructed (see Appendix C). Divided by the number of items, the scores on each scale ranged from 1 to 5. Reliability (Cronbach's alpha) was .78 for reading self-efficacy and .62 for reading task value.

## Results

Descriptive statistics and a correlation matrix for all variables are presented in Table 1. The seven variables were used in multiple regressions. Mean, standard deviation, skewness, kurtosis, and alpha are presented at the bottom of the table. As can be seen, coefficients of skewness ranged from  $-1.39$  to  $.46$ , and coefficients of kurtosis ranged from  $-.83$  to  $2.09$ . No score distribution was found to be substantially skewed, and they were considered suitable for use in parametric statistical analyses.

**TABLE 1** Descriptive Statistics and Correlations for All Variables

Variable	1	2	3	4	5	6	7
1. Word reading ability	—						
2. Listening comprehension	.15*	—					
3. Nonverbal ability	.25**	.27**	—				
4. Reading self-efficacy	.17*	.11	.10	—			
5. Reading task value	-.09	-.01	-.15*	.31**	—		
6. MC reading comprehension	.51**	.36**	.48**	.37**	-.20	—	.83**
7. CR reading comprehension	.48**	.40**	.39**	.29**	-.06	.83**	—
<i>M</i>	30.63	5.74	40.60	4.12	4.17	11.71	11.23
<i>SD</i>	10.06	1.62	8.01	.59	.50	4.87	4.6
Skewness	.46	-1.39	-1.14	-.90	-.81	-.36	-.49
Kurtosis	.17	1.25	2.09	.40	.63	-.83	-.51
Alpha	.86	.72		.78	.62	.86	.85

\* $p < .05$ . \*\* $p < .001$ .

### Whole Sample

*Do reading task value and reading self-efficacy make independent contributions to variance in reading comprehension scores over and above the contributions from decoding, listening comprehension, and non verbal ability in the MC and CR formats, respectively?*

Hierarchical multiple regression analysis was used. In the first step, word reading ability, listening comprehension, and nonverbal ability were entered. In the second step, reading self-efficacy and reading task value were added. This was done separately for the MC comprehension score and the CR comprehension score as dependent variables.

Table 2 shows the results of regression analyses of both MC and CR reading comprehension scores for the first and second steps. As regards MC items, the table shows that word reading ability ( $\beta = .40$ ,  $p = .000$ ), listening comprehension ( $\beta = .21$ ,  $p = .000$ ), and nonverbal ability ( $\beta = .33$ ,  $p = .000$ ) were all positive predictors of reading comprehension scores.<sup>1</sup> After controlling for the variance associated with the variables entered in the first step, the reading motivation variables accounted for additional variance. However, a statistically significant positive relationship

**TABLE 2** Results of Hierarchical Regression Analysis for Variables Predicting MC and CR Reading Comprehension

	MC Reading Comprehension <sup>a</sup>			CR Reading Comprehension <sup>b</sup>		
	<i>B</i>	<i>SE B</i>	$\beta$	<i>B</i>	<i>SE B</i>	$\beta$
Step 1						
Word reading ability	.19	.03	.40**	.17	.03	.38**
Listening comprehension	.62	.16	.21**	.80	.16	.28**
Nonverbal ability	.20	.03	.33**	.13	.03	.22**
Step 2						
Word reading ability	.17	.03	.36**	.16	.03	.34**
Listening comprehension	.56	.15	.19**	.76	.16	.27**
Nonverbal ability	.19	.03	.31**	.12	.03	.21**
Reading self-efficacy	2.18	.44	.26**	1.55	.45	.20*
Reading task value	-.21	.51	-.02	-.48	.52	-.05

*Note.* <sup>a</sup> $R^2 = .50$ ; for step 1  $F$  change = 54.51 (.000), for step 2  $F$  change = 13.17 ( $p = .000$ ).

<sup>b</sup> $R^2 = .41$ ; for step 1  $F$  change = 43.05 (.000), for step 2  $F$  change = 5.87 ( $p = .003$ ).

\* $p < .01$ . \*\* $p < .001$ .

was found only for reading self-efficacy ( $\beta = .26$ ,  $p = .000$ ), indicating that participants who believed themselves capable of doing well on reading comprehension tasks were more likely to perform well on MC reading comprehension questions.

As regards CR items, too, word reading ability ( $\beta = .38$ ,  $p = .000$ ), listening comprehension ( $\beta = .28$ ,  $p = .000$ ), and nonverbal ability ( $\beta = .22$ ,  $p = .000$ ) were positive predictors of reading comprehension scores. After controlling for the variance associated with the variables entered in the first step, the reading motivation variables accounted for additional variance in the CR format as well. Again, a statistically significant positive relationship was found only for reading self-efficacy ( $\beta = .20$ ,  $p = .001$ ), indicating that participants who believed themselves capable of doing well on reading comprehension tasks were more likely to perform well on CR reading comprehension questions as well.

As can be seen from Table 2, word reading ability accounted for a similar amount of variance for both item formats. Nonverbal ability and reading self-efficacy accounted for slightly more unique variance in MC scores than in CR scores.

Listening comprehension accounted for slightly more unique variance in CR scores than in MC scores.

The answer to the first research question was consequently that reading efficacy made independent contribution to variance in reading comprehension scores over and above the contributions from step one in both the MC and CR formats, whereas reading task value did not.

### *Groups With Different Levels of Reading Self-Efficacy*

*Do reading task value and/or reading self-efficacy make similar independent contributions, if any, to variance in reading comprehension scores for students at various levels of reading self-efficacy?*

To examine the second research question, the sample was split in two: students with low reading self-efficacy ( $N = 102$ ) and students with high reading self-efficacy ( $N = 115$ ). In the group with high reading self-efficacy (HRE) there were 52.2% boys and 47.8% girls. In the group with low reading self-efficacy (LRE) there were 45.1% boys and 54.9% girls. Table 3 provides results from independent  $t$ -tests and shows that there were no significant differences between these two groups in word reading ability,

**TABLE 3** Independent Sample  $t$ -Tests

Variable	Independent Sample $t$ -Test	Eta Squared
Word reading ability	No significant difference between HRE ( $M = 31.75, SD = 9.53$ ) and LRE ( $M = 29.37, SD = 10.54$ ); $t(217) = 1.74, p = .083$	.010
Listening comprehension	No significant difference between HRE ( $M = 5.84, SD = 1.58$ ) and LRE ( $M = 5.62, SD = 1.67$ ) $t(217) = .99, p = .322$	.004
Nonverbal ability	No significant difference between HRE ( $M = 41.08, SD = 8.38$ ) and LRE ( $M = 40.07, SD = 7.58$ ); $t(217) = .93, p = .355$	.004
MC reading comprehension	Significant difference between HRE ( $M = 13.01, SD = 4.47$ ) and LRE ( $M = 10.24, SD = 4.90$ ); $t(217) = 4.36, p = .000$	.100
CR reading comprehension	Significant difference between HRE ( $M = 12.17, SD = 4.42$ ) and LRE ( $M = 10.18, SD = 4.60$ ); $t(217) = 3.25, p = .001$	.046

listening comprehension, or nonverbal ability. However, there were significant differences, in favor of the HRE group, on both reading comprehension measures. The magnitude of the difference was small to moderate for CR comprehension scores but moderate to large for MC comprehension scores.

The analyses described in the previous section were then repeated separately for the LRE and HRE groups. Again, word reading ability, listening comprehension, and nonverbal ability were entered in the first step, and reading self-efficacy and reading task value were added in the second step.

Table 4 provides the results of the regression analysis for each of the groups on MC and CR comprehension, respectively. For HRE students, word reading ability ( $\beta = .33, p = .000$ ), listening comprehension ( $\beta = .27, p = .000$ ), and nonverbal ability ( $\beta = .40, p = .000$ ) were all positive predictors of MC reading comprehension scores. Word reading ability ( $\beta = .33, p = .000$ ), listening comprehension ( $\beta = .36, p = .000$ ), and nonverbal ability ( $\beta = .21, p = .014$ ) were also positive predictors of CR reading comprehension scores. After controlling for variance associated with the variables entered in the first step, the reading motivation variables did not account for any additional variance in either item format.

As regards LRE students, Table 4 shows that only word reading ability ( $\beta = .47, p = .000$ ) and nonverbal ability ( $\beta = .24, p = .004$ ) were significant positive predictors of MC reading comprehension scores in the first step. In the CR format word reading ability ( $\beta = .42, p = .000$ ), nonverbal ability ( $\beta = .21, p = .012$ ), and listening comprehension ( $\beta = .21, p = .013$ ) were all positive predictors. After controlling for variance associated with the variables entered in the first step, reading self-efficacy was found to account for additional variance ( $\beta = .23, p = .005$ ) in the MC format but not in the CR format.

The same pattern that emerged for the sample as a whole concerning the relative importance of the different predictors was also found when the sample was split into groups based on level of reading self-efficacy. Word reading ability accounted for a comparable amount of variance in both item formats. Nonverbal ability accounted for more unique variance in MC scores than in CR scores. Listening comprehension accounted for more unique variance in CR scores than in MC scores. However, there were differences between the two groups in the amount of unique variance



**TABLE 4** Results of Hierarchical Regression Analysis for Variables Predicting MC and CR Reading Comprehension for Students with Low and High Reading Self-Efficacy

	MC Reading Comprehension						CR Reading Comprehension					
	Students With High Reading Self-Efficacy <sup>a</sup>			Students With Low Reading Self-Efficacy <sup>b</sup>			Students With High Reading Self-Efficacy <sup>c</sup>			Students With Low Reading Self-Efficacy <sup>d</sup>		
	<i>B</i>	<i>SE B</i>	$\beta$	<i>B</i>	<i>SE B</i>	$\beta$	<i>B</i>	<i>SE B</i>	$\beta$	<i>B</i>	<i>SE B</i>	$\beta$
Step 1												
Word reading	.15	.03	.33***	.23	.04	.47***	.15	.04	.33***	.19	.04	.42***
Listening comprehension	.78	.21	.27***	.35	.24	.12	1.00	.23	.36***	.58	.23	.21*
Nonverbal ability	.22	.04	.40***	.16	.05	.24**	.11	.04	.21*	.13	.05	.21*
Step 2												
Word reading	.16	.03	.34***	.19	.04	.41***	.16	.04	.34***	.17	.04	.38***
Listening comprehension	.76	.21	.27***	.33	.23	.11	.98	.23	.35***	.56	.23	.20*
Nonverbal ability	.22	.04	.41***	.14	.05	.22**	.11	.05	.21*	.12	.05	.19*
Reading self-efficacy	.80	1.39	.04	2.39	.84	.23**	1.26	1.53	.06	1.56	.81	.16
Reading task value	.48	.66	.05	-1.12	.80	-.11	.19	.73	.02	-1.27	.78	-.14

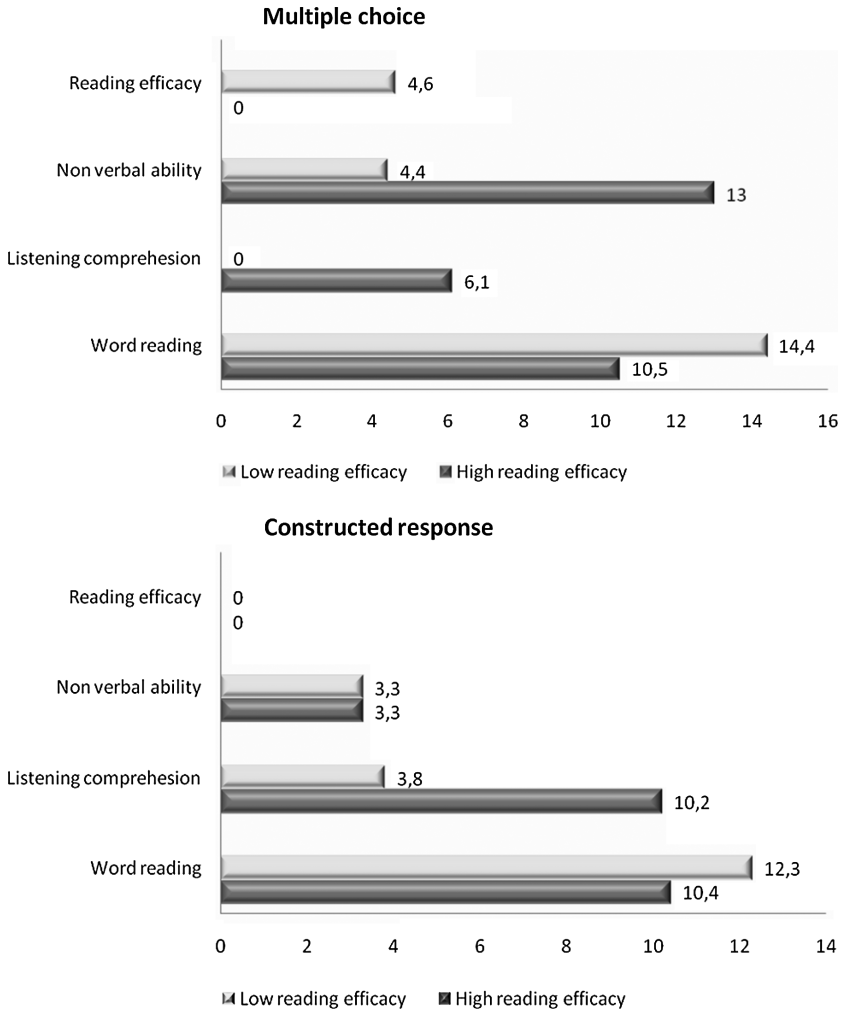
Note. <sup>a</sup> $R^2 = .51$ ; for step 1  $F$  change = 38.36 ( $p = .000$ ), for step 2  $F$  change = 0.41 (NS).

<sup>b</sup> $R^2 = .38$ ; for step 1  $F$  change = 19.75 ( $p = .000$ ), for step 2  $F$  change = 4.52 ( $p = .013$ ).

<sup>c</sup> $R^2 = .39$ ; for step 1  $F$  change = 23.14 ( $p = .000$ ), for step 2  $F$  change = 0.36 (NS).

<sup>d</sup> $R^2 = .36$ ; for step 1  $F$  change = 18.49 ( $p = .000$ ), for step 2  $F$  change = 2.76 (NS).

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



**FIGURE 2** Variance uniquely predicted (squared part correlations) by variables that significantly predict reading comprehension in terms of multiple-choice and constructed-response comprehension scores for students with high and low reading self-efficacy.

accounted for by the different variables. As Figure 2 shows, word reading ability accounted for more unique variance in comprehension scores for LRE students than for HRE students. Both listening comprehension and nonverbal ability were stronger predictors for HRE students than for LRE students.

Reading self-efficacy was not a unique predictor of CR comprehension scores when the sample was split—reading self-efficacy accounted for unique variance only in the case of MC comprehension scores for LRE students. For MC comprehension scores, reading self-efficacy was a stronger predictor than listening comprehension for LRE students.

### **Discussion and Consequences for Test Development**

The discussion of the results is divided into two parts. The first one discusses predictors of comprehension scores in HRE and LRE students, respectively. The second deals with the implication that multiple-choice questions will magnify the impact of self-efficacy in assessments of reading comprehension.

#### *Predictors of Reading Comprehension*

For HRE students, there was stability in terms of the variables predicting reading comprehension. Word reading ability, listening comprehension, and the ability for nonverbal reasoning all predicted scores on both MC and CR measures of reading comprehension. Nonverbal ability was a stronger predictor of MC scores than of CR scores, and listening comprehension was a stronger predictor of CR scores than of MC scores. These results support Allan's (1992) findings concerning strategies induced by the MC and CR format, respectively. If MC questions induce strategies that focus on the stem and the alternatives, including test-wiseness, then it is conceivable that students take advantage of their reasoning skills as measured by the Raven test. A conceivable consequence of CR questions inducing strategies that focus more on the reading passage and prior knowledge is that students' language skills will be more important for their success. HRE students seem to activate abilities that they will profit from while answering questions in each of the two item formats.

For LRE students, the picture was more complex. Word reading ability and nonverbal ability positively predicted reading comprehension scores on both measures. In addition, reading self-efficacy positively predicted MC scores, and listening comprehension positively predicted CR scores. In general, the variance uniquely predicted by listening comprehension and nonverbal

ability was lower for the LRE group than for the HRE group. This indicates that LRE students exploit these abilities less when answering questions about texts they have read; this finding would be in line with Bandura's (1997) claim that low-efficacious students may make poor use of their capabilities under certain circumstances.

*Do Multiple-Choice Questions Magnify the Impact of Self-Efficacy?*

It has been proposed that students with low self-efficacy in reading will struggle with complex reading tasks in an assessment situation, but what exactly is a "complex reading task?" Reading self-efficacy predicted MC comprehension scores but not CR comprehension scores for LRE students. This indicates that answering MC questions might be perceived as a more complex task for students with low reading self-efficacy than answering short-answer CR questions. The present study does not include any process measures or other information about students' thinking and/or behavior while answering questions. In the absence of process measures, two earlier studies of test-taking behavior will be highlighted in order to discuss why MC questions might be perceived as complicated for this group of students.

Langer (1987) used think-aloud methodology to explore how third graders solved MC questions in a reading test. The participants read a text and questions about it (MC comprehension questions without the response options) and were asked to provide answers. They were then shown the response options and asked to explain why they preferred one option over the others. Langer reported that students often selected incorrect options based on what she herself considered to be plausible interpretations of the text. She concluded that some students seemed to be disadvantaged by the MC format and were unable to demonstrate their understanding using the restricted options they were given. It is not at all surprising that a correct option in an MC question can be expressed in ways that are unexpected to children—after all, similar content can be expressed in a number of ways. This was, in fact, at one time an important argument in favor of the introduction of CR questions in large-scale assessments of reading (Valencia & Pearson, 1987), and it is taken into account in the scoring guides for CR test questions. But what exactly is it that a student

must do in order to decide whether a response option in an MC question has the expected semantic content? The student must probably weigh the various response options against each other and against the expected response and/or eliminate unlikely response options. This requires students to have the persistence to invest extra effort in order to demonstrate understanding. LRE students are less likely to do this if they find the task demanding.

Campbell (2005) presented profiles of the thinking processes used by two eighth graders answering MC and CR comprehension test questions. Campbell shows how a high-performing test taker seemed to “filter” his thinking about the text through options in an MC question: “[. . .] the options helped to scaffold and direct his thinking about the text” (Campbell, p. 362). An average-performing test taker, on the other hand, often selected options on MC questions without careful consideration. But when the average-performing student encountered CR questions, she was more likely to consider the text carefully while answering them. Campbell concluded that CR questions “force” interaction with the text in a way that MC questions do not and that the average-performing test taker profited from this.

In Langer’s (1987) study, some children lost score points on MC questions because they seemed unable to relate the answer they expected to the options offered. Campbell’s (2005) example has a different approach. Like Allan (1992), Campbell pointed out that the CR format, unlike the MC format, forces interaction with the text during answering and that some students profit from this. Both these arguments can help us to understand why the MC format can be a challenge for students with low reading self-efficacy. Both Langer and Campbell highlighted how the demonstration of understanding through the MC format can require persistence and effort to an extent that the format itself—given the presence of explicit response options—does not seem to encourage.

In a sample of ninth graders, Anmarkrud and Bråten (2009) found that even though the relationship between reading self-efficacy and reading task value, on the one hand, and reading comprehension (MC items), on the other, was the same in terms of simple correlations, reading self-efficacy was lost as a unique predictor when controlling for knowledge and strategy use. They

proposed that this was due to reading self-efficacy being more strongly related to achievement in the domain and to strategy use. It is possible that reading self-efficacy would have been lost as a unique predictor of MC comprehension scores in the present study, too, if strategy use had been controlled for. Not doing that, however, makes the distinctive feature of the MC format more obvious. Efficacy beliefs play a crucial role in decisions about how much effort to put into reading activities. The MC format probably induces problem-solving strategies that are more vulnerable to the influence of low reading self-efficacy than the strategies activated to deal with short-answer CR questions.

An additional explanation for the results beyond the extra energy required to process MC questions for LRE students is their instant reaction to the format of the question. For example, upon glancing at the MC format, an LRE student might be more likely to formulate an instant negative judgment about ability to handle such a question. LRE students may, however, perceive a CR response (with fewer words) as easier to read because of less verbiage. So, it may not be that LRE students fall back upon their prior experiences with MC questions to make judgments; it may be that they anticipate more work based on the visual format.

An important question is whether variance predicted by reading self-efficacy affects responses in a manner that is relevant or irrelevant to reading comprehension. According to Messick (1995), construct-irrelevant variance implies that an assessment is too broad: "... containing excess reliable variance associated with other distinct constructs as well as method variance such as response sets or guessing propensities that affects responses in a manner irrelevant to the interpreted construct" (p. 742). Is reading self-efficacy a fundamental aspect of comprehension, or is it a characteristic of the individual that affects comprehension but that cannot in itself be considered a fundamental aspect of comprehension?

We know that motivation affects the development of reading comprehension. Students with low self-efficacy seem to avoid challenging reading tasks, and by doing so they miss out on opportunities to improve their reading comprehension. By contrast, self-efficacious students choose demanding reading material and challenging reading tasks, which in turn influences their reading development in a positive way. This line of

argument views self-efficacy as a characteristic of the individual that affects the *development* of reading comprehension. The assumption that the connection between motivation and reading comprehension is mediated through strategy use is also based on a view of self-efficacy as a characteristic of the individual that affects reading comprehension, but *current achievement* is emphasized rather than development. When individuals believe they can be successful in an activity, they are more likely to engage in it (Bandura, 1997; Schunk & Zimmerman, 1997; Zimmerman, 1995). A reader who is active and engaged in the reading material will understand more from it than a passive reader (Guthrie & Wigfield, 2000; Pressley, 2002).

If the extent of active construction of meaning from text is a function of the level of self-efficacy, then variance uniquely predicted by reading self-efficacy is relevant to reading comprehension scores and should in principle affect reading comprehension scores in both item formats. Construct-irrelevant variance is systematic error that is group or person specific (Haladyna & Downing, 2004). The basic forms of construct-irrelevant variance are construct-irrelevant difficulty and construct-irrelevant easiness. *Construct-irrelevant difficulty* refers to contaminating influence that tends to systematically decrease test scores to some individuals or groups (Messick, 1995). The results of the present study show that differences in reading comprehension scores between students with high and low reading self-efficacy, respectively, are greater in MC reading comprehension than in short-answer CR reading comprehension and further that reading self-efficacy predicts MC reading comprehension for low-efficacious students. These findings indicate that the level of reading self-efficacy is more decisive to achieving points on MC questions than on short-answer CR questions—or, put in another way: that MC questions magnify the impact of self-efficacy for some students in assessments of reading comprehension.

The level of self-efficacy affects how much students understand of the texts they read but probably also the degree to which they are able to demonstrate what they have actually understood. Considering the results of the present study, we cannot rule out the possibility that students with low reading self-efficacy who fail MC questions would have been able to answer the same questions correctly if they had been designed as short-answer CR questions. If that is so, then the variance on MC questions predicted by

reading self-efficacy for students with low reading self-efficacy can be viewed as irrelevant method variance. For such conclusions to be drawn, however, the findings from the present study would have to be validated through other studies, and additional process studies of test-taking behaviors would have to be performed. In earlier studies we have used eye-tracking methodology to explore test-taking behaviors (Solheim & Uppstad, 2010). A natural next step in order to understand how students with different levels of reading self-efficacy cope with comprehension questions in different item formats will be to combine eye-tracking and think-aloud methodology.

### Note

1. In all the analyses there were only small differences in variance associated with word reading, listening comprehension, and nonverbal ability from step 1 to step 2. The  $\beta$  values given for these variables in the text refer to step 1.

### References

- Alderson, J. C. (2000). *Assessing reading*. Cambridge, UK: Cambridge University Press.
- Alexander, P. A. (2005). The path to competence: A lifespan developmental perspective on reading. *Journal of Literacy Research*, 37, 413–436.
- Allan, A. I. C. G. (1992). *EFL reading comprehension test validation: Investigating aspects of process approaches*. Unpublished PhD thesis, Lancaster University, Lancaster, England.
- Anmarkrud, Ø., & Bråten, I. (2009). Motivation for reading comprehension. *Learning and Individual Differences*, 19, 252–256.
- Baker, L., & Wigfield, L. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly*, 34, 452–477.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: Freeman.
- Campbell, J. (2005). Single instruments, multiple measures: Considering the use of multiple item formats to assess reading comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 347–368). Lawrence Erlbaum, Mahwah, NJ.
- Carr, M., Mizelle, N. B., & Charak, D. (1998). Motivation to read and learn from text. In C. R. Hynd & S. A. Stahl (Eds.), *Learning from text across conceptual domains* (pp. 45–70). Mahwah, NJ: Lawrence Erlbaum.
- Dermitzaki, I., Andreou, G., & Paraskeva, V. (2008). High and low reading comprehension achievers strategic behaviours and their relation to performance in a reading comprehension situation. *Reading Psychology*, 29, 471–492.
- Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109–132.



- Farr, R., Pritchard, R., & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. *Journal of Educational Measurement*, 27, 209–226.
- Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading*, 10, 323–330.
- Francis, D. J., Feltcher, J. M., Catts, H. W., & Tomblin, J. B. (2005). Dimensions affecting the assessment of reading comprehension. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 369–394). Lawrence Erlbaum, Mahwah, NJ.
- Guthrie, J. T., Hoa, A. L. W., Wigfield, A., Tonks, S. M., Humenick, N. M., & Littles, E. (2007). Reading motivation and reading comprehension growth in the later elementary years. *Contemporary Educational Psychology*, 32, 282–313.
- Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 403–424). Lawrence Erlbaum, Mahwah, NJ.
- Guthrie, J. T., & Wigfield, A. (2005). Roles of motivation and engagement in reading comprehension assessment. In S. G. Paris & S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 187–214). Lawrence Erlbaum, Mahwah, NJ.
- Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading*, 3, 231–256.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23, 17–27.
- Høien, T., & Tønnesen, G. (1998). *Håndbok til ordkjedetesten* [Handbook for the word-chain test]. Stavanger, Norway: Stiftelsen Dysleksisforskning.
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, 2, 127–160.
- Johnson, E. S., Jenkins, J. R., & Jewell, M. (2005). Analyzing components of performance assessment: An expanded simple view. *Reading Psychology*, 26, 267–283.
- Langer, J. (1987). The construction of meaning and the assessment of comprehension: An analysis of reader performance on standardized test items. In R. O. Freedle & R. P. Duran (Eds.), *Cognitive and linguistic analyses of text performance* (pp. 225–244). Norwood, NJ: Ablex.
- Messick, S. (1995). Validity of psychological assessment. Validation of inferences from persons' responses and performance as scientific inquiry into score meaning. *American Psychologist*, 50, 741–749.
- Mucherah, W., & Yoder, A. (2008). Motivation for reading and middle school students' performance on standardized testing in reading. *Reading Psychology*, 29, 214–235.
- Mullis, I. V. S., Kennedy, A. M., Martin, M. O., & Sainsbury, M. (2006). *PIRLS 2006 assessment framework and specifications* (2nd ed.). Chestnut Hill, MA: Boston College.
- Nasjonalt Læremiddelsenter. (1995). *Kartlegging av leseferdighet, 6. trinn* [Assessment of reading ability, Grade 6]. Oslo, Norway: Nasjonalt Læremiddelsenter.
- Pearson, P. D., & Hamm, D. N. (2005). The assessment of reading comprehension: A review of practices—Past, present and future. In S. G. Paris &

- S. A. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 13–69). Lawrence Erlbaum, Mahwah, NJ.
- Pintrich, P. (2000). An achievement goal perspective on issues in motivation terminology, theory and research. *Contemporary Educational Psychology*, *25*, 92–104.
- Pressley, M. (2002). Metacognition and self-regulated comprehension. In A. E. Farstrup & S. J. Samuels (Eds.), *What research has to say about reading instruction* (3rd ed., pp. 291–309). Newark, DE: International Reading Association.
- Raven, J. C. (1958). *Standard progressive matrices: Sets A, B, C, D & E*. Oxford, UK: Oxford Psychologists Press.
- Raven, J. C., Court, J. H., & Raven, J. (1988). *Standard progressive matrices*. London, UK: Oxford Psychologists Press.
- Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, *23*, 441–474.
- Schunk, D. H., & Zimmerman, B. J. (1997). Developing self-efficacious readers and writers: The role of social and self-regulatory processes. In J. T. Guthrie & A. Wigfield (Eds.), *Reading engagement: Motivating readers through integrated instruction* (pp. 34–50). Newark, DE: International Reading Association.
- Solheim, O. J., & Uppstad, P. H. (2007, August/September). *Reading during answering*. Paper presented at the 12th Biennial Conference for Research on Learning and Instruction, Budapest, Hungary.
- Solheim, O. J., & Uppstad, P. H. (2010). Eye tracking as a tool in process-oriented reading test validation. In O. J. Solheim (Ed.), *Vurdering av leseprestasjoner. Grunnleggende forutsetninger for måling av leseferdighet* [Assessment of reading performance. Fundamental conditions for measuring reading ability]. Unpublished doctoral dissertation, University of Stavanger, Stavanger, Norway.
- Valencia, S. W., & Pearson, P. D. (1987). Reading assessment. Time for a change. *The Reading Teacher*, *40*, 726–732.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy—Value theory of achievement motivation. *Contemporary Educational Psychology*, *25*, 68–81.
- Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation of reading to the amount and breadth of their reading. *Journal of Educational Psychology*, *89*, 420–432.
- Zimmerman, B. J. (1995). Self-efficacy and educational development. In A. Bandura (Ed.), *Self-efficacy in changing societies* (pp. 202–231). Cambridge, UK: Cambridge University Press.
- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, *25*, 82–91.

## Appendix A

### *Sample Items for the MC Reading Comprehension Measure*

NONFICTION: PROCESSES 1 AND 2

What is the Greek word for *knowledge*?

1. History
2. Politics
3. Archaeology
4. Logos

NONFICTION: PROCESSES 3 AND 4

“They taste extra good with Norwegian goat cheese and jam,” the text says. Which description fits this paragraph best?

1. Serving recommendations
2. Instructions
3. Advertisement
4. Recipe

FICTION: PROCESSES 1 AND 2

What does the lion father think is the worst that can happen to a lion?

1. Being eaten by a crocodile
2. Not being able to roar
3. Being laughed at
4. Only being able to eat mice

FICTION: PROCESSES 3 AND 4

What is the similarity between Karsten and Stian?

1. They are both proud and solemn
2. They both live in the jungle
3. They both have to leave their families but come home in the end
4. They are both laughed at, but in the end they are both admired

## Appendix B

### *Sample Items for the CR Reading Comprehension Measure*

NONFICTION: PROCESSES 1 AND 2

What animals pulled Medea's chariot?

NONFICTION: PROCESSES 3 AND 4

Why do you think that you ought not to fly the kite nearby a road that is carrying a lot of traffic?

FICTION: PROCESSES 1 AND 2

At the end of the story, Karsten's family no longer laugh at his squeaky roar. What has happened to make them stop laughing at Karsten?

FICTION: PROCESSES 3 AND 4

Why does Stian stick his front paw in his mouth to throw up what he has eaten?

## Appendix C

### *Items Used in Motivation Scales*

#### READING SELF-EFFICACY

1. I'm a good reader.
2. If a text is interesting, I don't care how hard it is to read.
3. It's easy for me to understand the content of a book.
4. I won't have any problems understanding what's in the textbooks in sixth grade.
5. Compared with the others in my class I'm a good reader.
6. If a book interests me, I read it even if it's long.
7. I think I'm doing well in reading this year.
8. I know that I'll do well on the reading comprehension tests in sixth grade.

#### READING TASK VALUE

9. I always think it's important to understand what I read.
10. Being a good reader is useful in all subjects.
11. Being a good reader is important to me.
12. I often talk to my friends about the things I read.
13. Being a good reader is useful when I have to learn new things.
14. Being a good reader is important to get a good job.
15. I don't think it's possible to do well at school without being good at reading.

Items 1, 2, 6, 8, 11, and 12 were adapted from Wigfield and Guthrie (1997).

Items 3, 5, 7, 9, 10, 14, and 15 were adapted from Anmarkrud and Bråten (2009).