



Reproduction Rather than Comprehension? Analysis of Gains in Students' Science Text Comprehension

Sascha Bernholt¹ · Hendrik Härtig² · Jan Retelsdorf³

Accepted: 4 August 2022 / Published online: 24 August 2022
© The Author(s) 2022

Abstract

The use of texts is an indispensable resource for students' learning, especially in science domains. While developing understanding of a specific topic usually is the main goal of reading expository texts, an important consideration is how to best measure whether this understanding has been reached. In this study, we aimed to analyze gains in students' reading comprehension based on reading three expository texts on chemistry and physics topics. By means of a pre-post design, we assessed the reading comprehension of 261 eighth grade students with regard to three levels of reading comprehension. Latent change scores were estimated to analyze changes in students' total test scores, while also calculating difference scores based on the single items. Results indicate that students' topic-related comprehension increases from pre- to posttest, while gains seem to be limited to word and sentence level questions. In line with other studies, these findings stress that students would benefit from explicit strategy instruction, at least when learning from reading is the goal of using science texts in classrooms.

Keywords Reading comprehension · reading assessment · expository texts · reading-to-learn

Introduction

A central instructional resource to support the acquisition of knowledge is the use of texts. This is especially true in the science domains, where the subject matter gets increasingly abstract and complex so that detailed verbal descriptions are indispensable (Pearson et al.,

✉ Sascha Bernholt
bernholt@leibniz-ipn.de

Hendrik Härtig
hendrik.haertig@uni-due.de

Jan Retelsdorf
jan.retelsdorf@uni-hamburg.de

¹ Leibniz Institute for Science and Mathematics Education (IPN), Olshausenstr. 62, 24118 Kiel, Germany

² Physics Education, University of Duisburg-Essen, Universitätsstr. 2, 45117 Essen, Germany

³ Faculty of Education, University of Hamburg, Von-Melle-Park 8, 20146 Hamburg, Germany

2010; Phillips & Norris, 2009; van den Broek, 2010). While developing understanding of a specific topic usually is the main goal of reading expository texts, an important consideration is how best to measure whether this understanding has been reached or not (Ahmed et al., 2016). Typically, summative assessments, often in the form of topic or content area tests (McCarthy et al., 2018), are employed after reading to ask readers to demonstrate their comprehension of the text. However, the question of gains in students' comprehension by reading a text is more difficult to answer (Beker et al., 2017). The present study aims to provide additional insights to this question by analyzing to what extent students' reading of three expository texts on chemistry and physics topics results in significant gains in comprehension as measured by a topic-related pre- and posttest and to what extent students' gains in comprehension differ between three levels of text comprehension, i.e., verbatim, propositional, and situation model.

Theoretical Background

Models of Text Comprehension

Different models of text comprehension can be found in the literature, each placing different emphases on specific cognitive processes that lead to the construction of a mental representation of a text (McNamara & Magliano, 2009). According to the construction-integration model (Kintsch, 1992; Van Dijk & Kintsch, 1992), reading can be conceptualized as the interaction between person and text. This interaction results in multiple levels of (mental) text representation that vary in terms of their abstraction in relation to the text at hand. Van Dijk and Kintsch (1992) distinguish between the three levels of words and phrases (i.e., the surface structure of single words and phrases as well as linguistic relations between them), semantic and rhetorical structure (i.e., relationships between sentences and sections of the text), and deep understanding. The highest level of understanding is reached when information provided by the text is elaborated and integrated with the reader's prior knowledge. This process results in a so-called situation model of the text (Kintsch, 1992; Schnotz, 2014), which consists of both inferences made from the text as well as information from the reader's knowledge base. In this context, the generation of inferences establishes connections between the currently read text and the previously read text or the background knowledge of the reader (Kraal et al., 2019).

Assessing Levels of Text Comprehension

With regard to comprehension outcomes, multiple approaches to conceptualize and to assess text comprehension can be found in the literature, ranging from rather superficial to deep understanding. The seminal works by van Dijk and Kintsch and the distinction between different levels of text comprehension have influenced a lot of research, albeit with different adaptations or labels. Schiefele (1999) distinguishes between the verbatim, propositional, and situational levels of text processing. However, topic-related prior knowledge (measured by factual multiple-choice items) was not significantly related to any component of text representation.

Ozuru et al. (2009) examined how text features (i.e., cohesion) and individual differences (i.e., reading skill and prior knowledge) contribute to text comprehension. By means of constructed-response items, the authors aimed to assess different levels of text

comprehension: text-based (requiring retrieval of “information explicitly stated within a sentence”); near- or local-bridging (requiring an “integration of information located within five clauses across multiple sentences (generally adjacent sentences)”); and far- or global-bridging (requiring “the integration of information located across larger distances, more than five clauses apart, and more than two sentences apart”; Ozuru et al., 2009, p. 230). Findings indicate that prior knowledge seems to play out stronger in global-bridging questions that require a more extensive integration of information (see Kintsch, 1992). With regard to text-based and local-bridging questions, the influence of general reading skill was larger, though still small compared to the effect of prior knowledge. The authors conclude that a deeper comprehension of a text (in terms of integrating the meanings of multiple sentences) is primarily determined by the knowledge participants already possess prior to reading (see Hwang, 2019; van den Broek, 2010).

In a similar approach, Ozuru et al. (2010) constructed open-ended questions to tap the three different levels of text-based, local-bridging, and global-bridging inferences. Results of an ANOVA revealed a main effect of the questions’ comprehension level, indicating that student scores decrease from text-based to global-bridging questions, even when taking the student’s prior knowledge and reading skill into account.

Learning and Comprehension

Generally, the process of creating a meaningful and coherent mental representation of a text during reading (i.e., text comprehension) and the process of encoding this information in long-term memory (i.e., learning) can be distinguished. Although text comprehension also involves retrieving and integrating information from long-term memory while building a mental representation of the text, learning is considered to reflect a relatively permanent change in the readers’ knowledge structure itself (van den Broek, 2010). With regard to different levels of text comprehension, it is often argued that “it is only the situation model that represents ‘real’ learning, whereas the verbatim and propositional representation reflect text memory or superficial learning” (Schaffner & Schiefele, 2007, p. 756).

As sketched above, creating a situation model of a text requires the reader to infer unstated or implied relationships based on their background knowledge and thus to integrate their knowledge with information extracted from the text. Consequently, background knowledge not only plays an essential role in most theories of text comprehension (McNamara & Magliano, 2009) but also in theories of learning (Bransford et al., 2000; Kendeou & O’Brien, 2016). In addition, empirical findings indicate that background knowledge impacts both textbase (Bransford & Johnson, 1972) as well as situation model understanding (McNamara & Kintsch, 1996). This is especially true when the new, text-based information contradicts the conceptions the reader already had before reading it (Penttinen et al., 2013).

In contrast to a view of reading as simple word recognition and information extraction, reading is also conceptualized as inquiry process (Norris & Phillips, 2008). According to this view, reading is a principled interpretation of the text. Deriving meanings from the text and integrating the textual information with relevant background knowledge corresponds to a search for meaning in the text in which justifications are provided, implied, and comprehensibly justified. Knowledge of a subject prior to reading is useless to a reader unless the reader recognizes the relevance of that knowledge by drawing inferences between the knowledge and the text (Phillips & Norris, 2009).

Overall, the generally assumed interrelation between text comprehension and learning is supported in numerous studies, by both theoretical arguments and empirical evidence. Also, the

major influence of prior knowledge on both text comprehension and learning is well grounded in the literature. However, the predominant use of overall scores for either background knowledge or for text comprehension (or both) in most studies makes it difficult to evaluate gains in text comprehension based on reading expository texts (Beker et al., 2017; McCarthy et al., 2018). This in turn stresses the importance of capturing different levels of text comprehension. However, levels of text comprehension are not independent of each other (Schnotz, 2014; Van Dijk & Kintsch, 1992). Establishing a situation model of the text requires both the text base as well as the reader's knowledge base. Thus, measures for different levels of text comprehension might be more or less indicative for changes in students' mental representation of a text during reading (i.e., text comprehension) (Gasparinatos & Grigoriadou, 2013).

The Present Study

The present study aims to provide additional insights by analyzing students' gains in comprehension based on reading expository science texts to answer the following research questions:

1. To what extent does students' reading of science texts result in significant gains in comprehension as measured by a topic-related pre- and posttest?
2. To what extent do students' gains in comprehension differ between the three levels of comprehension?

Based on the theoretical considerations sketched above, we expect significant gains in students' comprehension from pre to post based on reading different science texts. However, some measures are assumed to be more indicative of text memory (e.g., text-based questions), while other measures are assumed to be more sensitive to "real" learning" (Schaffner & Schiefele, 2007, p. 756). As multiple studies indicate that answering text-based questions requires less processing from the reader than answering inference questions (Gasparinatos & Grigoriadou, 2013; Ozuru et al., 2007), we expect smaller gains on higher levels of comprehension. However, significant gains should be noticeable on all three levels of text comprehension to be able to speak of a deeper understanding of the subject matter by reading the different texts, instead of a mere reproduction of text elements (Beker et al., 2017; Van Dijk & Kintsch, 1992).

Method

Sample

Data were collected in Germany in 2017. We draw on a sample of $N = 261$ students from grade eight (37.5% female; age: $M = 14.97$, $SD = 0.82$), attending different secondary schools in a larger metropolitan region. Participation was voluntary after the participants' caretakers provided written consent. Students were also informed that they could drop out of the study at any time without consequences.

We intentionally used introductory texts to different topics in the eighth grade which had not been taught to the students participating in this study. The underlying rationale was to reduce the variation in the students' prior knowledge, which is why we also expected the students to have little prior knowledge (McCarthy et al., 2018). A lower level of prior

knowledge was assumed to also ensure that there was at least some potential for comprehension gains and that students' would not be at the ceiling. However, the selected topics are curricular valid with regard to the students' grade level.

Measures

We decided to use introductory texts from science textbooks. Three texts were adapted from high school textbooks (one in the context of physics, two in the context of chemistry). The first text was about atomic physics, while the other two texts focused on acids and bases and chemical reactions, respectively. All three texts address curricular valid content and were not changed in terms of language. However, if applicable, graphic elements and possibly references to them were removed in order to focus on students' text comprehension. Linguistically, the texts can be characterized by a high information density and linguistic economy (complex prepositional phrases, nominalizations and attributions, resumption through abstracts and substitutions, multiple technical terms) as well as grammatical constructions typical for science texts (passive voice and impersonal sentence constructions). Texts were of about equal length (467 to 481 words) and in German language (see Supplemental material).

For each text, items were constructed and piloted that represent three levels of text comprehension according to typical frameworks (Kintsch, 1994; Ozuru et al., 2009; Schaffner & Schiefele, 2007; Schiefele, 1999). For the first type of question (verbatim), students were to extract information from single sentences. For the second type of question (propositional), students were to draw inferences between at least two sentences. For the third type of question (situation model), students were to integrate information from the text and their prior knowledge into a situational model of the text. For each text, 13 to 15 questions (42 items in total) were administered, identical in pre- and posttest and comprising both multiple-choice and constructed-response items (see Supplemental material for example items). The order of question levels within each test was randomized.¹ With regard to the sample of eighth grade students, we expected little prior knowledge to all three topics as these are usually taught later than our study took place.

Procedure

The pretest to measure students' topic-related knowledge prior to reading took place about 2 weeks before the students read the three science texts. After reading the texts, we assessed the students' topic-related knowledge of the three science texts again.

Analytical Strategy

Each student answered 42 items (13 to 15 per text; cf. Table 1) and students' answers were coded as correct (1) or incorrect (0) based on a predefined rubric. Data was analyzed by means of confirmatory factor analysis. First, we employed a one-factor model per measurement point (pre, post) to infer the models' fit to the data. After ensuring measurement

¹ Besides, general reading measures (general reading comprehension of narrative texts, vocabulary knowledge, inferences, and reading strategies) were administered to students, but results on these measures are beyond the scope of this manuscript.

Table 1 Descriptive statistics for the pre- and posttests

		MP	<i>n</i>	<i>Min</i>	<i>Max</i>	<i>M</i>	<i>SD</i>	McDonald's ω
Levels of comprehension	Level 1 – verbatim	pre	13	0	0.77	0.33	0.16	0.51
		post	13	0	1	0.5	0.21	0.72
	Level 2 – propositional	pre	13	0	0.85	0.32	0.16	0.50
		post	13	0.08	1	0.48	0.19	0.61
	Level 3 – situation model	pre	16	0	0.81	0.36	0.15	0.58
		post	16	0	0.88	0.39	0.17	0.66
Total	pre	42	0.02	0.76	0.34	0.13	0.76	
	post	42	0.12	0.88	0.45	0.16	0.84	

MP, measurement point; *n*, numer of items; *Min*, minimum; *Max*, maxium; *M*, mean; *SD*, standard deviation

invariance, we then examined whether students' performed differently over time (i.e., before and after reading) before researching indications of gains in text comprehension. For this purpose, we estimated univariate latent change score models (LCSM; McArdle, 2009) on students' pre- and posttests (reflecting changes in students' total test scores) to investigate students' overall gains in comprehension by means of structural-equation modeling (SEM). In this model, the difference between the two latent variables at two measurement points (i.e., students' pre- and posttest scores, respectively) is estimated as a third latent variable which represents the part of the posttest variable which is not identical to the pretest variable, i.e., the latent change variable mimics the result of a subtraction of the pretest score from the posttest score (cf. Fig. 1; McArdle, 2009).

Based on true individual change models (TICM; Steyer et al., 2014), which are a reparametrization of LCSMs, we again estimated the change in students' total test scores from pre to posttest, albeit with a focus on the mean structure (Fig. 2). In a TICM, all items (pre and post) load on a knowledge factor and post items additionally load on a gain factor. While factor loadings of identical items (pre/post) on the knowledge factor are restricted to be equal, factor loadings on the gain factor are freely estimated. In addition, residuals of identical items (pre/post) were allowed to correlate (correlated uniqueness approach). Based on this model, we also estimated differences in item thresholds between pre- and posttest for all items. These threshold differences reflect whether an item has become easier (i.e., solved more frequently; negative value) or harder (i.e., solved less frequently; positive value) from pre to post. These shifts in the difficulty of items were taken to indicate gains in comprehension based on reading the science texts.

While estimating the LCSM and the (structural model of the) TICM aims to provide results with regard to changes in students' total test scores from pre to post, the additional estimation of differences between item thresholds provides insights into changes of students' answer patterns with regard to single items. Estimating these differences while taking measurement error into account is a benefit of the SEM framework, as opposed to manifest approaches such as analysis of (co)variance (McArdle, 2009). Also, this approach makes it possible to estimate confidence intervals for these threshold differences to indicate the "security" of the students' gains in comprehension that are assumed to be reflected by these differences.

All SEM analyses were conducted in Mplus 8 using the robust weighted least squares mean and variance (WLSMV) estimator and theta parameterization. In addition, the correlated uniqueness approach and latent standardization method were employed (Little et al.,

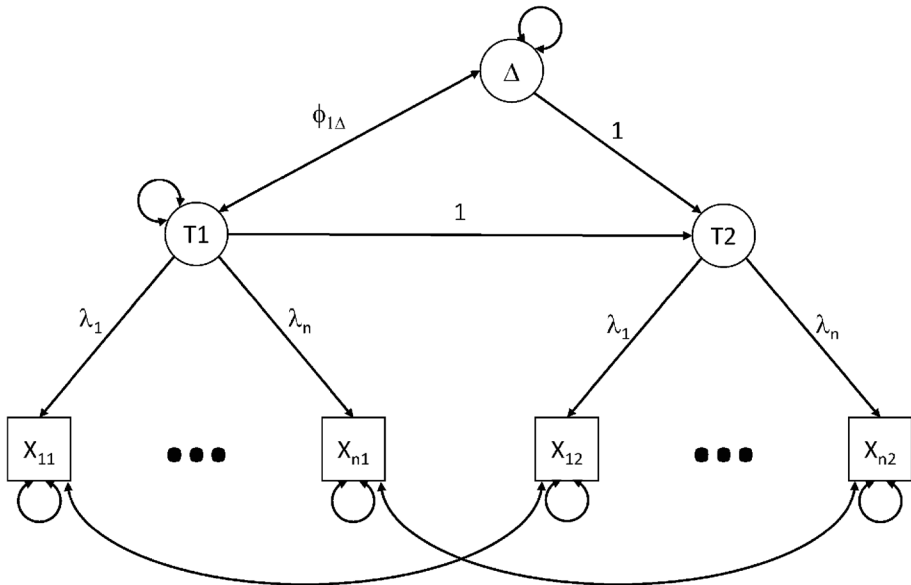


Fig. 1 Latent change score model with two latent factors ($T1$, $T2$) for the two measurement occasions (pre- and posttest) based on the same set of indicators (X_1 to X_n). Change between measurement occasions is modeled as an additional latent factor (Δ), including covariation of change with the initial factor ($\phi_{1\Delta}$). Factor loadings of identical items (pre/post) are restricted to be equal across measurement occasions. Residuals are allowed to correlate between measurement occasions

2007) to control for systematic measurement error and to scale the latent factors. Common thresholds were used for the comparative fit index ($CFI > .90$ indicating acceptable, $CFI > .95$ excellent fit) and the root-mean-square error of approximation ($RMSEA < .08$ indicating acceptable, $RMSEA < .06$ good fit) (Hu & Bentler, 1999).

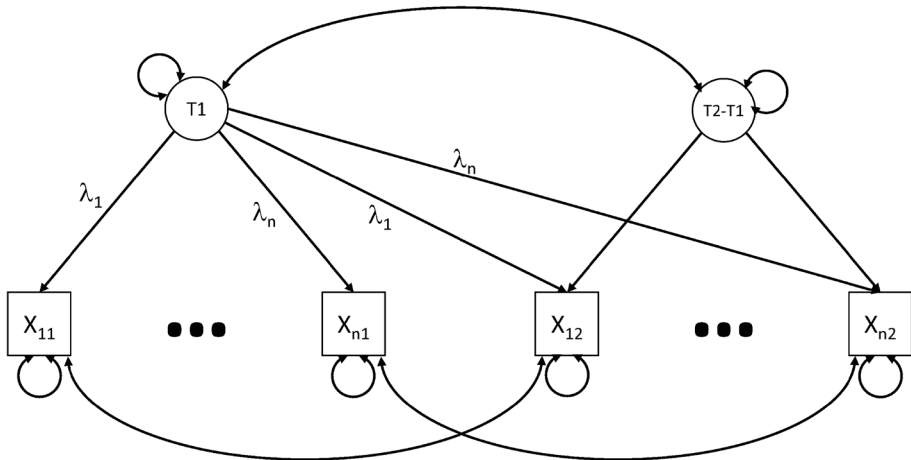


Fig. 2 True individual change model with two latent variables reflecting students’ pretest score ($T1$) and the true change score from pre- to posttest ($T2-T1$), respectively (cf. Steyer et al., 1997). Factor loadings of identical items (pre/post) on the first latent variable are restricted to be equal, factor loadings on the change factor are freely estimated. Residuals are allowed to correlate between measurement occasions

Reliabilities at each of the three levels of comprehension were below common thresholds, while the total test across all topic areas and levels of understanding features sufficiently high reliabilities (cf. Table 1). With regard to the students' total scores, descriptive results indicate low prior knowledge (total mean score of .34) and small gains from pre to post (total mean score of .45), but also substantial variation between students. Accordingly, some students solved hardly any items (indicated by a minimum value of 0), while some students answered all items correctly (in the posttest).

Results

As a first step, students' answers to pre- and post-tests were jointly modeled in a two-factor CFA model with acceptable fit indices ($CFI = .916$, $RMSEA = .016$, $WRMR = 1.005$). Intercorrelations between the two factors (i.e., measurement points) are high ($r = .91$). Composite reliabilities indicate high internal consistency for both factors (T1: .83; T2: .90)².

Gains in Overall Text Comprehension

We conducted a latent change score model (LCSM) to investigate students' gains in text comprehension from pre- to posttest. Based on partial strong measurement invariance (cf. Supplemental material), we strived to establish longitudinal structural equivalence to test for changes in latent means and for the impact of the pretest score on this change over time. The model indicates an adequate fit to the data ($CFI = .90$, $RMSEA = .017$, $WRMR = 1.062$).

The results of the LCSM indicate a latent mean value of $M_{pre} = 2.02$ ($SE = 0.38$) and a latent change of $M_{change} = 0.37$ ($SE = 0.05$). Moreover, the change of students' text comprehension from pre- to posttest seems to be uncorrelated with the pretest scores ($r = -0.04$, $p = .53$).

Gains in Text Comprehension on Different Levels of Comprehension

Based on a true individual change model (Steyer et al., 2014), differences of item thresholds between pre- and posttest were estimated. As a complement to this model, thresholds for all items (pre and post) were estimated and subtracted (post minus pre) to calculate a difference score that reflects whether an item has become easier (i.e., solved more frequently; negative value) or harder (i.e., solved less frequently; positive value) from pre to post. Also, 95% confidence intervals were estimated for these difference scores (cf. Fig. 3).

The model indicates a good fit to the data ($CFI = .914$, $RMSEA = .016$, $WRMR = 1.008$). Based on the estimated item thresholds, difference scores from pre to post were analyzed (cf. Table 2). On average, almost identical mean difference scores for verbatim (mean difference $M = -.53$, significantly different from 0 according to Wald test, $\chi^2(1) =$

² Three-dimensional models with separate factors per topic area partially provided a slightly better fit to the data. However, intercorrelations between the different topic areas within measurement points were high (latent correlations from CFA models [.65; .93]) and composite reliabilities of the single dimensions in the three-factor CFA models were substantially below common thresholds. Hence, the following analyses are conducted based on one-dimensional factor-models by combining the test results from different topic areas.

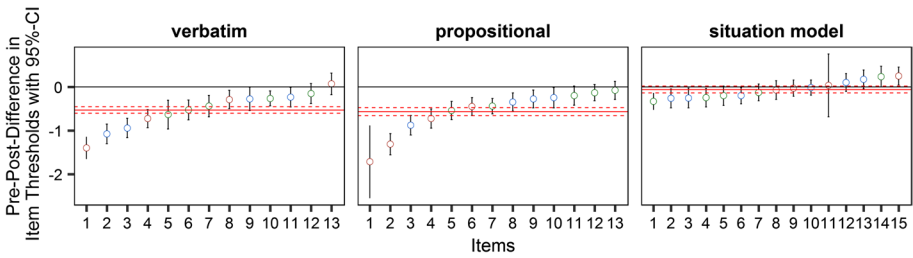


Fig. 3 Difference in item thresholds from pre to post and corresponding 95% confidence interval for all items of each of the three question levels (verbatim, propositional, and situation model) and topic areas (atomic structure – red; chemical reaction – blue; acids/bases – green)

Table 2 Mean difference scores for items addressing different levels of text comprehension

Topic area	Level 1 (verbatim)	Level 2 (propositional)	Level 3 (situation model)
Acids & bases	-0.63	-0.44	-0.07
Atomic structure	-0.58	-1.05	0.05
Chemical reaction	-0.40	-0.28	-0.13
Total	-0.53	-0.56	-0.06

192.08, $p < .001$) and propositional questions (mean difference $M = -.56$, significantly different from 0 according to Wald test, $\chi^2(1) = 146.83, p < .001$), while difference scores for situation model questions are close to zero (mean difference $M = -.06$, not significantly different from 0 according to Wald test, $\chi^2(1) = 2.09, p = .15$).

By means of ANOVA, we analyzed whether topic area and/or question level had an effect on students’ comprehension gains (cf. Table 3). The effect of comprehension level is significant ($F(2, 32) = 9.51, p < .001$) and can be considered as large (*partial* $\omega^2 = 0.30$). Both the effect of topic area ($F(2, 32) = 1.89, p = .17, \textit{partial} \omega^2 = 0.04$) and the interaction between comprehension level and topic area are not significant ($F(4, 32) = 2.51, p = 0.06, \textit{partial} \omega^2 = 0.13$).

While especially the factor of comprehension level explains a substantial amount of variance (*partial* $\eta^2 = .37, \textit{partial} \omega^2 = .29$), there is still substantial variation between items within the three levels of text comprehension. Notably, variance is smallest for situation model items. Here, 95% confidence intervals for 10 out of 15 items indicate difference scores not significantly different from zero, while two items even indicate negative shifts from pre to posttest, i.e., students on average performed worse on these items after reading the texts.

Discussion

In this study, we aimed to analyze students’ gains in comprehension based on reading expository science texts. Based on overall test scores, the mean values for students’ topic-related text comprehension indicate significant gains from pre to post. This is in line with other studies that applied topic-related knowledge tests (McCarthy et al., 2018; Ozuru et al., 2009). However, most of these studies report findings based on correlations or

Table 3 One-way analysis of variance of difference scores by comprehension level and topic area

	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η^2	<i>partial</i> η^2	ω^2	<i>partial</i> ω^2
Comprehension level	2	1.12	9.51	< .001	0.29	0.37	0.26	0.29
Topic area	2	0.22	1.89	.17	0.06	0.11	0.03	0.04
Interaction	4	0.30	2.51	.06	0.16	0.24	0.09	0.13
Residuals	32	0.12	–	–	–	–	–	–

MS, mean squares

regression, aiming at providing evidence for the effect of prior knowledge on students' text comprehension (Ahmed et al., 2016; Hwang, 2019). In addition, most studies conceptualize, on the one hand, students' prior knowledge in terms of a rather broad and more general understanding of the domain or topic and, on the other hand, their understanding of the specific content of the text, which is often seen as much more focused and fine-grained (McCarthy et al., 2018; Ozuru et al., 2009), making it difficult to compare the results reported here to other findings.

Desiron et al. (2018) assessed students' comprehension about two scientific phenomena (tsunami, drought) by using the same multiple-choice items pre and post. Students' answers were coded and then used to calculate their relative knowledge score to capture "the knowledge gain as a percentage of the potential learning gain" (Desiron et al., 2018, p. 472). When applying this approach to the results reported here, students' posttest scores were on average about 20% higher than their pretest scores (ranging 18 to 24% between topics), which is somewhat lower than the values of 28 to 38% reported in Desiron et al. (2018). While this discrepancy might be due to differences in item format and student age, both the study by Desiron et al. (2018) and the present study support the claim that reading expository texts results in "a significant increase in pupils' knowledge between pre-test and immediate post-test" (Desiron et al., 2018, p. 465).

With regard to changes in students' answers to single items, the conclusions with regard to gains in text comprehension become more pronounced. Students' overall gains seem to be predominantly based on word and sentence level questions, while questions on the level of the situation model indicate no gains.

According to Gasparinatou and Grigoriadou (2013), text-based questions are more indicative of text memory while inference questions are more sensitive to learning (see Schaffner & Schiefele, 2007) or comprehension (see Beker et al., 2017; Kintsch, 1992). From this perspective, students participating in this study may have memorized some information from reading the three texts, but they may not have learned a lot of new information (cf. Ozuru et al., 2009). As such, it is questionable whether students actually were able to develop a situation model of the text and to integrate text content to their own prior knowledge (Kintsch, 1988), which is considered necessary to be able to apply this information from the text to solve novel problems or to answer complex questions (Beker et al., 2017). As only the situation model is often considered to represent "real" learning" (Schaffner & Schiefele, 2007, p. 756), the conclusion here is rather that based on reading the three expository texts no learning or only superficial learning has actually taken place.

A possible reason might be, as indicated by observational classroom studies, that "science is taught through lecture, demonstrations, or textbooks that are designed to 'deliver content' to students rather than actively engaging students in the work of making sense of

science phenomena” (Goldman et al., 2016, p. 231). As a result, students are accustomed to searching scientific texts for information rather than engaging intellectually with texts to develop deeper understanding (Berland & Hammer, 2012), resulting in surface and text base level representations of the text content.

The low level of students’ prior knowledge (as discussed above) might have been the limiting resource in developing a deeper understanding (Hwang, 2019; McCarthy et al., 2018). As students were unfamiliar with the topics addressed in the different texts, they might have simply lacked substantial prior knowledge or they might have had problems in activating their prior knowledge relevant to the topics at hand. As expository texts tend to place higher processing demands on the reader, due to structural complexity, information density, or abstract and logical relations (Penttinen et al., 2013), students might not have had enough cognitive resources to elaborate on the information provided by the text and to integrate this information with their prior knowledge (Hwang, 2019).

Overall, the results presented here raise questions regarding the effects that the composition of tests and the analytical strategy that either picks up or ignores this composition have on conclusions regarding students’ comprehension from reading expository texts (McCarthy et al., 2018). Despite numerous studies in the field of text comprehension, it is still difficult to accurately diagnose a readers’ “true” comprehension of a text and how this comprehension changes over the course of reading. While the adequacy of common textbook texts to support students’ learning is frequently considered low (Härtig et al., 2019), findings from this and other studies stress that students would benefit from explicit strategy instruction (Dori et al., 2018; Jian, 2020; McNamara & Kintsch, 1996), at least when “our concern is not whether students recall the textbook but whether they learn from it.” (Kintsch, 1994, 295). However, engaging students to use science texts as a resource for scientific inquiry requires fundamental changes in science education (Goldman et al., 2016; Greenleaf & Valencia, 2017), pertaining to the perception of texts use in classroom by teachers, the quality of the texts available for science instruction, and both students’ and teachers’ literacy skills with regard to reading, writing, teaching, and learning from science texts (Pearson et al., 2010).

Limitations

As we intentionally used introductory texts in this study, the overall low scores in the tests as well as low reliabilities of the topic-specific tests may not be surprising. Students probably had only little prior knowledge that they could draw upon when both answering the test questions and when reading the three texts. This might limit the generalizability of the findings as results might differ when employing texts that are more closely related to students’ prior knowledge (Kintsch, 1994). However, both the selected texts and the selected topics are curricular valid with regard to the students’ grade level. It can therefore be assumed that the content of the texts was appropriate for the target group.

For the purpose of studying gains in comprehension through reading, we employed the same test before and after reading the different texts. Applying the same measure twice, however, might result in re-test effects due to sensitization or selective activation of elements in students’ prior knowledge by specific terms or questions in the pretest. As sensitization might influence students to focus on specific elements during reading which they might have missed without this indication by a pretest, this could be considered a threat to validity. Similarly, the pretest might have primed students to look for specific content during reading in order to understand these aspects. To minimize these effects, we included a delay of two weeks between pretesting and the actual reading experiment (cf. Swanborn &

de Glopper, 1999). When nevertheless assuming re-tests effects to be in place, the findings reported here seem to indicate that these effects seem to be limited to verbatim and propositional questions, while repeated administration of tests seems not to effect answering situation model questions. As the design of the present study does not allow differentiating the effects of re-testing and reading, additional studies might build up on the reported findings to analyze further, how the assessment influences inferences about the outcomes of students' reading.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11165-022-10066-6>.

Acknowledgments The research reported in this article was part of the project 'Comprehension of Scientific Texts as an Interaction between Reader and Text Characteristics', directed by Hendrik Härtig (University of Duisburg-Essen), Jan Retelsdorf (University of Hamburg) and Sascha Bernholt (Leibniz Institute for Science and Mathematics Education Kiel). We would like to thank Nicole Fraser for her huge contribution to the project as well as Gabriel Nagy for advice on the analytical strategy. The project was supported by the German Research Foundation (DFG; Project number 263507343).

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Competing Interests The authors have no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahmed, Y., Francis, D. J., York, M., Fletcher, J. M., Barnes, M., & Kulesz, P. (2016). Validation of the direct and inferential mediation (DIME) model of reading comprehension in grades 7 through 12. *Contemporary Educational Psychology, 44–45*, 68–82. <https://doi.org/10.1016/j.cedpsych.2016.02.002>
- Beker, K., Jolles, D., & van den Broek, P. (2017). Meaningful learning from texts: The construction of knowledge representations. In J. A. León & I. Escudero (Eds.), *Reading Comprehension in Educational Settings* (pp. 29–62). John Benjamins Publishing Company.
- Berland, L. K., & Hammer, D. (2012). Framing for scientific argumentation. *Journal of Research in Science Teaching, 49*(1), 68–94. <https://doi.org/10.1002/tea.20446>
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience and school*. National Academy Press.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 717–726. [https://doi.org/10.1016/S0022-5371\(72\)80006-9](https://doi.org/10.1016/S0022-5371(72)80006-9)
- Desiron, J. C., de Vries, E., Bartel, A. N., & Varahamurti, N. (2018). The influence of text cohesion and picture detail on young readers' knowledge of science topics. *British Journal of Educational Psychology, 88*(3), 465–479. <https://doi.org/10.1111/bjep.12195>

- Dori, Y. J., Avargil, S., Kohen, Z., & Saar, L. (2018). Context-based learning and metacognitive prompts for enhancing scientific text comprehension. *International Journal of Science Education*, 40(10), 1198–1220. <https://doi.org/10.1080/09500693.2018.1470351>
- Gasparinatou, A., & Grigoriadou, M. (2013). Exploring the effect of background knowledge and text cohesion on learning from texts in computer science. *Educational Psychology*, 33(6), 645–670. <https://doi.org/10.1080/01443410.2013.790309>
- Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., . . . READI, P. (2016). Disciplinary Literacies and Learning to Read for Understanding: A Conceptual Framework for Disciplinary Literacy. *Educational Psychologist*, 51(2), 219–246. <https://doi.org/10.1080/00461520.2016.1168741>
- Greenleaf, C., & Valencia, S. (2017). Missing in action: Learning from texts in subject-matter classrooms. In K. A. Hinchman & D. A. Appleman (Eds.), *Adolescent literacies: A handbook of practice-based research* (pp. 235–256). The Guilford Press.
- Härtig, H., Fraser, N., Bernholt, S., & Retelsdorf, J. (2019). Kann man Sachtexte vereinfachen? – Ergebnisse einer Generalisierungsstudie zum Textverständnis. *Zeitschrift für Didaktik der Naturwissenschaften*, 25(1), 273–287. <https://doi.org/10.1007/s40573-019-00105-7>
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hwang, H. (2019). The role of science domain knowledge and reading motivation in predicting informational and narrative reading comprehension in L1 and L2: An international study. *Learning and Individual Differences*, 76(2), 101782. <https://doi.org/10.1016/j.lindif.2019.101782>
- Jian, Y.-C. (2020). Teaching fourth-grade students of different reading abilities to read biological illustrations and integrate in-text information: An empirical experiment. *Research in Science Education*, 50(6), 2269–2282. <https://doi.org/10.1007/s11165-018-9778-8>
- Kendeou, P., & O'Brien, E. J. (2016). Prior knowledge: Acquisition and revision. In P. Afferbach (Ed.), *Handbook of individual differences in reading: Reader, text, and context* (pp. 151–163). Routledge Taylor and Francis Group.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 95(2), 163–182. <https://doi.org/10.1037/0033-295X.95.2.163>
- Kintsch, W. (1992). A cognitive architecture for comprehension. In H. L. Pick, P. W. van den Broek, & D. C. Knill (Eds.), *Cognition: Conceptual and methodological issues* (pp. 143–164). American Psychological Association.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist*, 49(294–303).
- Kraal, A., van den Broek, P. W., Koornneef, A. W., Ganushchak, L. Y., & Saab, N. (2019). Differences in text processing by low- and high-comprehending beginning readers of expository and narrative texts: Evidence from eye movements. *Learning and Individual Differences*, 74(4), 101752. <https://doi.org/10.1016/j.lindif.2019.101752>
- Little, T. D., Card, N. A., Slegers, D. W., & Ledford, E. C. (2007). Representing contextual effects in multiple-group MACS models. In T. D. Little, J. A. Bovaird, & N. A. Card (Eds.), *Modeling contextual effects in longitudinal studies* (pp. 121–147). Lawrence Erlbaum Associates.
- McArdle, J. J. (2009). Latent variable modeling of differences and changes with longitudinal data. *Annual Review of Psychology*, 60, 577–605. <https://doi.org/10.1146/annurev.psych.60.110707.163612>
- McCarthy, K. S., Guerrero, T. A., Kent, K. M., Allen, L. K., McNamara, D. S., Chao, S.-F., . . . Sabatini, J. (2018). Comprehension in a scenario-based assessment: Domain and topic-specific background knowledge. *Discourse Processes*, 55(5–6), 510–524. <https://doi.org/10.1080/0163853X.2018.1460159>
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, 22(3), 247–288. <https://doi.org/10.1080/01638539609544975>
- McNamara, D. S., & Magliano, J. P. (2009). Toward a comprehensive model of comprehension. In B. H. Ross (Ed.), *Psychology of Learning and Motivation*, 51(51), 297–384. [https://doi.org/10.1016/S0079-7421\(09\)51009-2](https://doi.org/10.1016/S0079-7421(09)51009-2)
- Norris, S. P., & Phillips, L. M. (2008). Reading as Inquiry. In R. A. Duschl & R. E. Grandy (Eds.), *Teaching Scientific Inquiry* (pp. 233–262). BRILL. https://doi.org/10.1163/9789460911453_018
- Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, 25(4), 399–438. <https://doi.org/10.1080/07370000701632371>
- Ozuru, Y., Briner, S., Best, R., & McNamara, D. S. (2010). Contributions of self-explanation to comprehension of high- and low-cohesion texts. *Discourse Processes*, 47(8), 641–667. <https://doi.org/10.1080/01638531003628809>

- Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction, 19*(3), 228–242. <https://doi.org/10.1016/j.learninstruc.2008.04.003>
- Pearson, P. D., Moje, E., & Greenleaf, C. (2010). Literacy and science: Each in the service of the other. *Science (New York, N.Y.), 328*(5977), 459–463. <https://doi.org/10.1126/science.1182595>
- Penttinen, M., Anto, E., & Mikkilä-Erdmann, M. (2013). Conceptual change, text comprehension and eye movements during reading. *Research in Science Education, 43*(4), 1407–1434. <https://doi.org/10.1007/s11165-012-9313-2>
- Phillips, L. M., & Norris, S. P. (2009). Bridging the gap between the language of science and the language of school science through the use of adapted primary literature. *Research in Science Education, 39*(3), 313–319. <https://doi.org/10.1007/s11165-008-9111-z>
- Schaffner, E., & Schiefele, U. (2007). The effect of experimental manipulation of student motivation on the situational representation of text. *Learning and Instruction, 17*(6), 755–772. <https://doi.org/10.1016/j.learninstruc.2007.09.015>
- Schiefele, U. (1999). Interest and learning from text. *Scientific Studies of Reading, 3*(3), 257–279. https://doi.org/10.1207/s1532799xssr0303_4
- Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 72–103). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.006>
- Steyer, R., Partchev, I., & Shanahan, M. J. (2014). Modeling true intraindividual change in structural equation models: The case of poverty and children’s psychological adjustment. In T. D. Little, K. U. Schnabel, & J. Baumert (Eds.), *Modeling longitudinal and multilevel data: Practical issues, applied approaches, and specific examples*. Psychology Press.
- Swanborn, M. S. L., & de Glopper, K. (1999). Incidental word learning while reading: A meta-analysis. *Review of Educational Research, 69*(3), 261–285.
- Van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science, 328*(5977), 453–456. <https://doi.org/10.1126/science.1182594>
- Van Dijk, T. A., & Kintsch, W. (1992). *Strategies of discourse comprehension*. Academic.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.