**Instructed highlighting of text passages – Indicator of reading or strategic performance?**

Nora Heyne[1]: nora.heyne@uni-bamberg.de

Cordula Artelt[1]: cordula.artelt@lifbi.de

Timo Gnambs[1, 2]: timo.gnambs@lifbi.de

Karin Gehrer[1]: karin.gehrer@lifbi.de

Cornelia Schoor[3]: cornelia.schoor@uni-bamberg.de

Affiliations:

[1]Leibniz Institute for Educational Trajectories, Wilhelmsplatz 3, 96047 Bamberg, Germany;

[2]Institute for Education and Psychology, Johannes Kepler University Linz, Altenberger Str. 69, 4040 Linz, Austria;

[3]Faculty of Human Sciences and Education, Department of Educational Research, University of Bamberg, Markusplatz 3, 96047 Bamberg, Germany.

Corresponding author: Nora Heyne

Business address: Faculty of Human Sciences and Education, Center of Teacher Education, University of Bamberg, Luisenstr. 5, 96047 Bamberg, Germany, nora.heyne@uni-bamberg.de, telephone: 0049 951 863 3182 (manuscript was prepared while contract at affiliation[1])

Private address: Maria-Ward-Str. 186, 96047 Bamberg, Germany, N_Heyne@web.de, mobile: 0049 176 23844433.

Publication and Licence

Heyne, N., Artelt, C., Gnambs, T., Gehrer, K., & Schoor, C. (2020). Instructed highlighting of text passages – Indicator of reading or strategic performance? Lingua, 236. https://doi.org/10.1016/j.lingua.2020.102803



Funding:

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of interest:

none

Acknowledgements:

Our cordial thank goes to Florian Kopp, Dr. Ilka Wolter as well as to all of the colleagues at the Leibniz Institute for Educational Trajectories in Bamberg (Germany) and associated colleagues who work to implement NEPS and thus enabled this study. Furthermore, we thank Amanda Habbershaw for proofreading.

# 1 Introduction

*Highlighting* while reading and learning is a very common procedure. People of different ages use it when reading for different purposes and in various situations. In particular, they mark information in texts during reading in order to find or remember it easily for subsequent learning or further processing or to meet other more or less explicitly defined goals. Usually this technique is introduced to pupils in primary school reading classrooms for the first time (Artelt, 2006; Heyne, 2015) and is used most frequently later on by older pupils and students while reading and learning (e.g., Kobayashi, 2009; Ponce & Mayer, 2014). As consistently found in several studies, more than 80 percent of college students regularly use highlighting (Peterson, 1992). This has been corroborated by self-reports and objective observations of student behaviour while learning (Brennan, Winograd, Bridge & Hiebert, 1986).

In contrast to this self-paced use of highlighting, which is primarily performed spontaneously in the context of self-regulated learning, *instructed highlighting* refers to the selective marking of text passages following a predefined instruction or in order to answer a given question about a specific text. This sort of situation usually emerges during instruction in classrooms of young pupils or might be used in administrations of competence tests. From a theoretical point of view, instructed highlighting is expected to require various cognitive operations that characterize reading and learning strategies. Therefore, in this study within the National Educational Panel Study (NEPS; Blossfeld & von Maurice, 2011)[1] and within the frame of developing competence tests, we assumed that instructed highlighting partly represents capabilities of reading as

[1]**Abbreviations:** NEPS = National Educational Panel Study; *P* = Precision; *O* = Overplus; *C* = Correspondence with experts; *D* = Divergence from experts; *MS* = Metacognitive strategies; *CS* = Cognitive strategies; *MR* = Management of resources; *RE* = reading competence; WLE = weighted maximum likelihood estimates; LS = learning strategies; LIST = questionnaire *Lernstrategien im Studium*.

well as strategic performance. Depending on the capabilities which are indicated by instructed highlighting empirically, this sort of task could be used for technology-based assessments of reading or strategic performance in future.

In this context, one aim of this study was to describe highlighting behaviour by means of objective and across different texts valid measures, in particular by means of introduced indices. Furthermore, it was aimed to determine how these indices are related to reading competence (RE) and the use of learning strategies (LS), from a theoretical and empirical point of view. Finally, results were expected to indicate if items of instructed highlighting within a technology-based assessment might be diagnostically conclusive for measuring reading competence or strategic performance.

## 2 Theoretical framework

From a theoretical perspective and from the point of view of a cognitive task analysis, instructed highlighting is assumed to require diverse operations, such as 1) keeping the question in mind, 2) reading and understanding the text – in particular, decoding symbols and words, assigning them to semantic issues, constructing a coherent text representation, inferring conclusions, interpreting and reflecting on the text (e.g., Kintsch, 1998; Lenhard & Artelt, 2009; Gehrer et al., 2013) – and, 3) checking and selecting text passages with reference to the question, and finally 4) marking or not marking these text passages. According to literature, highlighting generally supports initial encoding, superficial processing of texts and rote memorization (Mayer, 1996; Ponce & Mayer, 2014; Peterson, 1992). Furthermore, it can improve information selection, acquisition and transfer into the working memory, in particular when an individual has to master complex learning tasks (Weinstein & Mayer, 1986). According to Leopold and Leutner (2015), this selection function of highlighting only effects information processing if important text passages are highlighted selectively. However, learners seldom exclusively focus on selected aspects, tend to mark (too) many passages and show pronounced individual differences in the use of highlighting (e.g., Leopold & Leutner, 2015; Peterson, 1992). Beyond these studies on highlighting in reading of specific texts, there are few approaches to describe highlighting behavior and to investigate its relations to reading competence and to strategic performance which is in focus of the study.

## 2.1 A rationale for coding highlighting

As noticed above, highlighting is a very popular learning strategy when reading, which is performed in various ways. Therefore, differences of quality have to be considered when describing highlighting behaviour. In particular in order to assess instructed highlighting that has the objective of answering a given question, various quality aspects should be distinguished with reference to the master solution. By analogy with the differentiation of detecting signals in the signal detection theory (Green & Swets, 1966), four types of results can be distinguished when highlighting text passages in order to answer a specific question. Within the signal detection theory, the following four cases are distinguished: 1) the *Hit*, if a specific signal was given and also detected; 2) the *Correct Rejection*, if no specific signal was given and also not detected; 3) the *Missing*, if a given signal was not detected and 4) the *False Alarm*, if no signal was given, but a signal was detected. With reference to this distinction, four cases are distinguished to describe instructed highlighting in this study. Thus, instructed highlighting results can entail 1) highlighting of solution-relevant elements (hit), 2) no highlighting of non-relevant elements (correct rejection), 3) highlighting of non-relevant elements (false alarm) and 4) no highlighting of solution-relevant elements (missing).

Based on these cases, ideal instructed highlighting should exclusively be applied to solution-relevant parts without any additional highlighting. Therefore, all non-relevant text passages should not be highlighted, but only those parts that are relevant for the solution. Hence, all four cases should be taken into account to describe instructed highlighting, which is the focus of the first question of this study. Futhermore, we aimed to find measures which are objective, valid and applicable across different texts.

Recent literature has not provided any methods for assessing highlighting results that incorporate these cases and criteria. Instead, systems for scoring selected features of highlighting results or sequence analyses of highlighted symbols are usually applied. For example, in the study of Leopold and Leutner (2015), highlighting of main ideas, important concepts and non-relevant information in texts was scored based on precursory ratings of experts. Whereas this scoring system might have been beneficial for investigating the differences in highlighting between participants working on the same texts, it is not expected to be conclusive for analysing instructed highlighting across texts, which is the objective of this study. Furthermore, comparisons of highlighted sequences of symbols were made in order to find the longest common subsequences by means of sequence analytical methods (Sukkarieh, von Davier, & Yamato, 2012). These also did not reveal information on the amount of correctly or falsely (not) highlighted symbols, which is the objective of this study. In addition, questionnaires in which participants estimate their own highlighting behaviour are not expected to provide the information sought. Moreover, their use could also give rise to problems. As stated in the literature, self-reports on the use of strategies by students might not necessarily show what students do when studying (e.g., Artelt, 2000; Winne & Jamieson-Noel, 2002; Veenman, 2005). Nevertheless, this deviation could pertain only to specific types of strategies or learning aids and not be generalized. Brennan and colleagues (1986) found a high level of agreement between self-reports and observations on the use of highlighting when undergraduate students, mainly sophomores ($N$ = 50), were reading and summarizing a factual text (1800 word excerpt). By contrast, Brennan and colleagues found lower correspondence between reported and observed strategies for note-taking, repeated reading and other aspects. However, because we found no appropriate measures in recent literature for evaluating quality aspects of instructed highlighting with the chosen focus, we derived four indices in this study, which are introduced in the following part (see Table 1). By contrast to existing scoring systems, analytical approaches or questionnaires for self-reports, we expected the proposed parameters to provide rather objective, reliable and valid values in order to capture all four cases described above. Furthermore, by taking text lengths and the amount of solution-relevant and non-relevant parts at symbol level into account, they allow comparisons across different texts.

Table 1

*Denotation and calculation of quality indexes of highlighting*

| Description | Calculation | With: |
|---|---|---|
| Precision ($P$) | $P = r / R$ | $r$: Number of correctly highlighted signs[*1]<br>$R$: Number of correct signs |
| Overplus ($O$) | $O = f / F$ | $f$: Number of incorrectly highlighted signs<br>$F$: Number of incorrect signs |
| Divergence from experts´ judgements ($D$)[*2] | $D = \sum_{k=1}^{n} \lvert HL_{\mathrm{S}} - HL_{\mathrm{EX}} \rvert \; x \; T_{\mathrm{i}} / T_{\mathrm{t}}$ | $HL_S$: Subjects´ judgements<br>$HL_{EX}$: Experts´ judgements<br>$T_i$: Sign number of text unit<br>$T_t$: Sign number of text in total |
| Correspondence with experts´ judgements ($C$) | $C = \frac{Z_{00} + Z_{11} - (Z_{01} + Z_{10})}{n}$ | $Z_{11}$. $Z_{00}$: Number of concordant judgements<br>$Z_{10}$. $Z_{01}$: Number of divergent judgements<br>$n$: Number of cases (text units) |

*Note.* [*1] Correctness is assigned when judgements of subjects are similar to experts´ judgements; [*2] by contrast to $C$, $D$ is weighted for text units´ lenghts.

First of all, the index *Precision* (*P)* was defined as a parameter for how exactly individuals find elements in a text that are the right solution to a given question according to experts' views. This value corresponds to Case 1 mentioned above (cf. hit) and indicates the amount of correctly chosen parts of a text in relation to the parts of the text that have to be chosen according to the experts. Theoretically, the value can range from $P = 0$ which means no part of the right solution has been marked to $P = 1$, which means that the right passages have been highlighted completely.

The index *Overplus* ($O$) indicates the amount of additional text passages that has been highlighted but that was not part of the correct answer. This value is calculated based on the amount of falsely marked words relative to the total amount of solution-irrelevant text passages. It thus corresponds to Case 4 (false alarm) with reference to the signal detection theory. Based on this index, the results can range from $O = 0$ which means that none of the non-relevant text passages have been marked to highlighting of all of the text elements that are not relevant for the answer based on the experts´ view ($O = 1$).

The index *divergence from experts´ judgements* ($D$) provides a measure of discrepancies between the test takers´ and the experts´ judgements. This value is calculated by the sum of the differences at the level of text units, which are weighted by the text units' lengths respectively. Results for this value can show no discrepancy ($D = 0$) through to a maximum amount of divergence between participants´ and experts´ judgements ($D = 1$). Hence, all incorrect judgements of persons (cf. false alarms and missings) are integrated in this index. By contrast to the following index $C$, the advantage of index $D$ is that it takes into account the lengths of text units.

The index *correspondence* ($C$) with experts´ judgements provides a measure of the amount of accordance between the participants´ and experts´ solutions. Based on the measure of correspondence of Holley and Guilford (1964) and formulas proposed by Eckert (1998), it is calculated as the sum of all concordant answers ($Z_{11}$, $Z_{00}$) from which all divergent answers ($Z_{10}$, $Z_{01}$) are subtracted, weighted by the number of all cases and text units ($n$) respectively. By calculating this value, all false and correct judgements of the test takers are integrated into one parameter and it therefore provides a holistic measure of all four mentioned cases taking into account the text units. Values for $C$ can show perfect concordance between the participants´ and experts´ judgements ($C$ = 1) through to no concordance with respect to all text units ($C$ = -1).

## 2.2 Relation of instructed highlighting to reading competence

As mentioned above, we assume that instructed highlighting requires different processes of reading, i.e., phonetic recoding, the identification of words, activation of semantic meanings and recoding to recognize the text passages that represent the answer to a given question. By contrast to these basal reading operations, instructed highlighting can also demand more complex reading processes such as constructing coherence at sentence level or across the entire text (e.g., Kintsch, 1998; Lenhard & Artelt, 2009); individuals are also required to make inferences, interpretations and evaluate the text. Beside these processes and depending on the predefined instruction, instructed highlighting might also demand a) the identification of single facts, b) inference of conclusions or c) evaluation and interpretation, which characterize operations of reading competence in the frame conception of the study (Gehrer et al., 2013). Therefore, this study addresses the relation between reading competence and results of instructed highlighting.

## 2.3 Relation of instructed highlighting to the use of strategies

Instructed highlighting also entails various operations that are conducted when using learning strategies. Therefore, capabilities in using learning strategies are expected to correlate with instructed highlighting. Learning strategies are defined as ”any thoughts, behaviours, beliefs or emotions that facilitate the acquisition, understanding, or later transfer of new knowledge and skills” (Weinstein, Husman, & Dierking, 2000, p. 727). Furthermore, they are characterized as goal-orientated, consciously conducted and metacognitively controlled learning techniques (e.g., Artelt, 2000). The literature largely agrees on a distinction between cognitive, metacognitive and resource-related strategies, which support different processes of learning and are closely linked with each other (e.g., Weinstein & Mayer, 1986; Pintrich & Garcia, 1994; Leopold, den Elzen-Rump, & Leutner, 2006). For example, cognitive strategies support immediate information processing. As one type of cognitive strategy, rehearsal facilitates the selection of relevant information (Weinstein & Mayer, 1986; Wild & Schiefele, 1994) and can be enhanced by means of study aids requiring low transformation of text contents such as highlighting and note-taking (Wild, 2006). This selecting and focusing on specific information requires concentration and, hence, the management of internal resources as one type of resource-related strategy. As an additional type of cognitive strategy, elaboration involves the integration of information from texts and previous knowledge, text-related inferences as well as reflection on what the text is about (Wild & Schiefele, 1994). Moreover, metacognitive strategies involve planning, monitoring and regulating reading and learning processes. For example, monitoring entails checking text understanding with respect to a given question, whereas regulation is characterized by initiating or changing learning behaviours such as individuals reading text passages again if they do not understand them (Friedrich & Mandl, 1997; Pintrich & Garcia, 1994).

According to the results of the Programme for International Student Assessment

(PISA), in particular controlling and elaborating are highly correlated with reading competence in almost all participating countries (Artelt et al., 2001). Furthermore, the experimental study of Leopold and colleagues (2006) showed that the use and, in particular, the quality of learning strategies of pupils of different grade levels correlated significantly with their learning outcomes when they read expository texts. In a subsequent study by Leopold and Leutner (2015) on the effects of highlighting trainings with and without fostering metacognitive regulation and monitoring, no significant differences were found in the text comprehension of participants in both conditions. In contrast, the experimental group with trained metacognitive abilities highlighted less unimportant information and, hence, selected information in a more sophisticated manner. Leopold and Leutner (2915) reported, that learners seldom exclusively focus on selected aspects but tend to mark too many passages. Furthermore, they found pronounced individual differences in the amount of additional highlighting beyond what was necessary for the solution; it varied between 18 and 862 words for a text with 1,470 words overall, in a percentage of 1 to 58 percent. In line with these results, most students (about 54 %) reported highlighting a third to half of a text; less than 9 percent of the participants highlighted between 0-10 percent of a text (Peterson, 1992). Thus, students obviously do not use highlighting selectively. Based on these reported results and approaches, strong relations were expected between instructed highlighting and the use of learning strategies, in particular, regulation, concentration and elaboration.

# 3 Research Questions

Against this theoretical background, the investigation of quality aspects of instructed highlighting and its relations to reading competence and the use of learning strategies is in focus of this study. Hence, the first question of the study involves providing a detailed description of instructed highlighting with respect to different aspects of quality, in particular, precision, overplus as well as divergence from and correspondence with experts' judgements. According to the above-mentioned results on the features of highlighting, we expected most test takers to highlight too much and, therefore, show high overplus, which was, however, expected to vary to a large degree between individuals. This additional highlighting with reference to experts´ solutions should also be indicated in the divergence from experts´ judgements (Parameter *D*), which was expected to reach positive values.

Since instructed highlighting involves numerous operations which are also needed for reading, highlighting capabilities were expected to correlate with reading competence to a high degree. Based on the reported assumptions and results, readers with good test results were expected to highlight with high precision and low overplus and to highlight the passages in agreement with the experts´ highlighting. This is because they decode, understand and construct text coherence fluently and, hence, easily recognize the answer to the question. In contrast, persons with low reading test results were assumed to highlight less precisely with higher overplus and also to highlight passages that deviate to a large extent from the ideal highlighting solution, which is indicated by a high divergence from but low correspondence with experts´ judgements.

The third question focuses on the relations between quality aspects of instructed highlighting and the use of learning strategies. Because concentration involves activities that facilitate attention, which are important for focusing on and selecting specific information from texts, it was expected to correlate strongly with instructed highlighting. Because monitoring and regulation are necessary for selecting and highlighting information in texts, and highlighting results are better after training of these capabilities (Leopold & Leutner, 2015), strong associations were expected between both metacognitive strategies and instructed highlighting. Thus, the higher a person's capabilities of regulation and monitoring are, the lower his or her expected values for overplus are. In addition, a lower divergence was expected from the experts´ solutions, and a person's highlighting was expected to be with a larger degree of correspondence with experts´ judgements. As mentioned above, elaboration involves information integration, reflection and inferences about the text and goes beyond superficial text features, which are in focus when highlighting. Therefore, lower correlations were expected between elaboration and instructed highlighting.

# 4 Methods

## 4.1 Sample and Design

The sample of the study includes *N* = 937 participants ($M_{age}$ = 31.9) from all over Germany (50.7% from big cities with more than 500,000 inhabitants; 13.6% from local communities and smaller cities to 50,000 inhabitants and 35.7% between), who were recruited by a professional survey institute, partly via the registration offices. About half of the sample (*n* = 443; female 50.8 %, $M_{age}$ = 25.6) involved students from universities (*n* = 3) and technical colleges (*n* = 4) across different disciplines, e.g., law, economics, social sciences, linguistics, cultural studies, mathematics, natural science, engineering, humanities, art and others. The other half (*n* = 494; female 55.3%, $M_{age}$ = 38.01) consisted of adults in three age groups, 18 to 25 years (36.4%), 26 to 45 years (30.4%) and 46 to 70 years (33.2%), stratified by educational status. In this educational stratification, adults with low education, e.g., with general education school leaving certificate obtained on completion of grade 9 at the German *Hauptschule*, were in the first group (33.7%; female 53.9%; $M_{age}$ = 36.38). Adults with middle education, e.g., general education school leaving certificate obtained on completion of grade 10 at German *Realschulen*, were in the second group (33%) (female 57.1%; $M_{age}$ = 38.06 ).

The third group (33.3%; female 54.4%; $M_{age}$ = 39.61) entailed adults with a general higher education entrance qualification. All participants were tested individually in their private homes via computerized assessments. First, they received standardized tests for the assessment of instructed highlighting and reading competence. These tests included three modules with different items (processing time about 28 minutes each) which were administered with a rotated multi-matrix design to control for order effects. Finally, each respondent filled in a computerized self-report questionnaire.

## 4.2 Measures

*Instructed Highlighting*. Highlighting tasks were implemented via computer-based assessment. Each task consisted of one text and one question which had to be answered by highlighting text passages that represented the answer a) as single information, b) as a basis for inferences or c) as a starting point for evaluation or reflection (see definition of reading competence above). The questions referred to one or more passages of the text read. Slightly changed examples of these questions from a development study are a) "What was typical for that time?" (as requirement of the extraction of information), b) "What must be given, so that the described process leads to a positive result?" (as requirement of drawing conclusions) and c) "Which text passage(s) indicate(s) the basic idea of the statement of person xy?" (as example of a reflective question). The participant has to decide which and how many texts passages to mark and over how many words or symbols the marking of the respective text passages would be continued usefully. Since all the symbols and words were available within the text, this task can be considered as an approximation to an open format. Altogether, six texts with questions were prepared as highlighting tasks (for an example, see Figure A.1 in the Appendix). The preparation of coding the answers of these tasks involved dividing the highlighting-task texts into units (with various numbers of symbols; $M = 37.57$; $SD = 31.03$) due to limits in the storage capacity of the system used. Following this, a variable was generated for each of these units in the dataset. Assigned to these variables, the answers of the test takers were scored by the computer as highlighted completely (1.0), partially (0.5) or not at all (0.0) and saved in files. Preceding the evaluation of these data, the relevance of the text units with respect to the given question was defined (relevant = 1; not relevant = 0). With reference to this *master solution*, the indices of quality aspects of instructed highlighting were calculated based on the participants´ codes: *Precision* (*P*), *Overplus* (*O*), *Divergence* (*D*) from and *Correspondence* (*C*) with experts´ judgements. With these indices, we aimed to describe how precisely test takers highlighted the passages of texts which were considered as relevant (Precision) and to which extent they underlined additional passages (Overplus). Furthermore, the indices *D* and *C* provided more general measures for the concordance with and divergence from experts´ judgements respectively.

*Reading competence.* Reading competence was captured by means of a computer-based Rasch-scaled test which included 20 texts. Based on the frame conception of reading competence in the NEPS, the test contained 147 questions with three levels of cognitive requirements (a. identification of single facts, b. inference of conclusions, c. evaluation and interpretation) with four different answering formats. The majority of the answering formats were multiple-choice items ($N = 72$), the others were complex items with several subtasks, like decision tables ($N = 35$), matching items ($N = 18$) and text-enrichment-tasks ($N = 16$) (Gehrer et. al, 2013; Weinert et al., 2011). During the course of data analysis, the subtasks of all complex items except from multiple-choice items were aggregated to produce items with partially correct solutions (for the scoring of the partial credit items; see Pohl & Carstensen, 2013). The test scores were given as weighted maximum likelihood estimates (WLE) and ranged from about -3.74 to 4.90 logits with a mean of 0.05 and a standard deviation of 1.13; they exhibited satisfactory reliability (.98), which was estimated based on all items of the pool. The parameters of the scale and the items showed a satisfactory model fit. The descriptive data from the application of this test are shown in Table 4 below.

*Learning strategies.* The use of learning strategies was measured by means of selected items from the questionnaire *Lernstrategien im Studium* (LIST; Wild & Schiefele, 1994) which assesses the use of learning strategies within the course of studies. Learning strategies are defined as effective and flexible activities chosen to optimize learning results. Referring to the *Motivated Strategies for Learning Questionnaire* (Pintrich, Smith, Garcia, & McKeachie, 1991) and the *Learning and Study Strategies Inventory* (Weinstein, Schulte, & Palmer, 1988), the LIST distinguishes between several cognitive, metacognitive and resource-related strategies (Wild & Schiefele, 1994, p. 186). Metacognitive strategies are primarily conceptualized as comprehension-monitoring strategies, including planning, monitoring and regulation. As a type of resource-related strategy, the management of internal resources includes attention management which involves consciously allocating concentration when learning.

Table 2

*Scales and questionnaire items of the use of learning strategies*

| Construct | Scale | Items |
|---|---|---|
| Learning strategies (LS) | Regulation (MS) | 1.1 Exactly rereading of not understood passages of texts |
| | | 1.2 Concentrated reading if text is complex |
| | | 1.3 Repeatedly reading if text is not understood |
| | Elaboration (CS) | 2.1 Referring to own experiences while reading |
| | | 2.2 Searching for concrete examples while reading |
| | | 2.3 Generating pictorial imaginations while reading |
| | Concentration (MR) | 3.1 Digressing throughts from subject while reading (r) |
| | | 3.2 Easy distraction while reading (r) |
| | | 3.3 Being mentally absent while reading (r) |

*Note.* Items of the use of learning strategies are taken out of the LIST (Wild & Schiefele, 1994); r = reversed polarity.

In order to perform a scenario-based assessment of learning strategies in this study, the items of the three scales were selected from the LIST because they referred to reading in the recent test situation. As shown in Table 2, three items were taken out of the *Elaboration* scale as one type of cognitive strategy (CS); they refer to the integration and connection of elements within the text and also with reference to previous knowledge (e.g., „I related what I read to my own experiences."). Another three items were used from the *Regulation* scale as a metacognitive and comprehension monitoring strategy (MS). This is defined as adapting reading activities due to monitoring of results, in particular, slowing down and rereading a text if it is not understood (e.g., "If I did not understand a text when I first read it, I went through it again step by step"). Furthermore, three items from the *Concentration* scale, as one type of management of internal resources (MR), were used in this study (e.g., "As I read, I caught myself wandering off with my thoughts."). All items were about activities of individual learning and had to be answered on a 4-point frequency scale (1 = very seldom, rarely, rather, 4 = very often). The empirical values in this study covered the full range of the scale and stretched from 1 to 4. Regarding elaboration and regulation, the means were mostly located centrally with moderate standard deviations; that is, participants reported using these strategies moderately often. Both constructs had acceptable reliability (α = .66). In contrast, concentration had a high reliability (α = .82). Because its general mean was slightly over the central point of the scale, strategies of concentration were used slightly more often than elaboration and regulation when learning. As the analyses of manifest intercorrelations show, the learning strategies did not correlate with each other, with the exception of elaboration and regulation ($r$ = .11; $p$ = .00). However, there was no significant manifest correlation of concentration with elaboration or regulation strategies respectively. Descriptive data on the application of these scales are shown in Table 4 below.

### 4.3 Statistical procedures

The structure of the quality aspects of highlighting was examined using confirmatory factor analyses in Mplus 7 (Muthen & Muthen, 2012). In particular, the indices Precision, Overplus, Divergence from and Correspondence with experts´ judgements were modeled as latent factors with all values from the different texts as indicators (cf. Table.1 in Appendix). The reliabilities were quantified by means of McDonalds Omega (McNeish, 2018). The focal hypotheses on the relations of instructed highlighting to reading competence and to using learning strategies were examined via correlation analyses based on structural equation models in Mplus 7.

## 5 Results

### 5.1 Description of instructed highlighting by means of quality indices

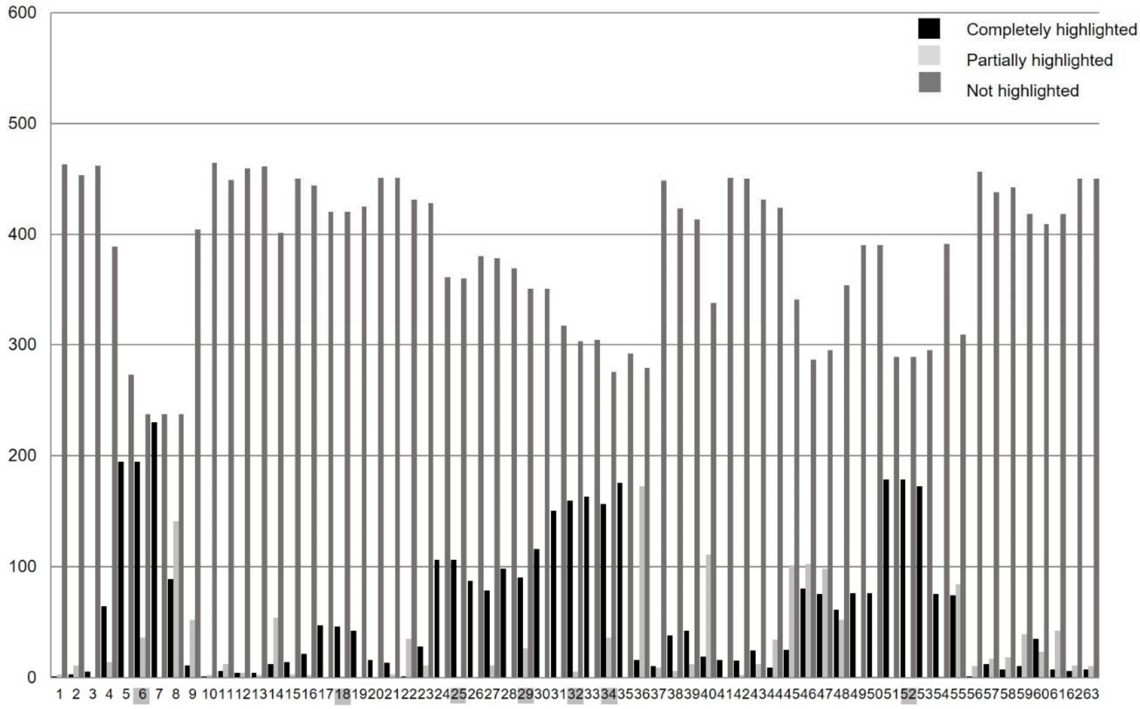


Figure 1. Instructed highlighting of subjects in Text 2 (correct units by experts´ solutions indicated by grey boxes; $N_{Missings}$ = 470)

A first insight into the highlighting results is presented in Figure 1. The units of Text 2 are displayed along the x-axis and shaded with grey boxes if they were correct in accordance with experts´ judgements. The bars with different grey scales indicate the frequencies of test takers with different solutions. The black bars show the numbers of participants who marked the text units completely, and the light-grey bars represent the frequencies of individuals who highlighted the text units partially. The dark-grey bars show the numbers of test takers who did not highlight the text units. The remaining participants of the sample did not work on this highlighting task by design (cf. rotated multi-matrix design) or due to non-compliant test behavior. As indicated in the picture, complete highlighting (black bars) occurred moderately often and most frequently for solution-relevant text units and neighboring text passages. The high frequencies of the dark-grey bars indicate that many individuals did not highlight the assigned text units at all. As visualized by the light-grey bars, partial highlighting occurred the least. Furthermore, all of the text units were highlighted at least by a few of the subjects either completely or partially, and no text unit was not marked at all. Similar patterns were found for all of the five other texts (see also Figure A.2 –A.6 in the Appendix).

The calculated indices of the participants´ highlighting in all texts are presented in Table 3. The values for precision (*P)* have very low means and moderate standard deviations but cover the full range of the scale with the exception of highlighting in Text 5, where only a maximum of .92 was achieved. As an index for more highlighting than necessary for the right solution, overplus (*O*) ranged from .00 to .73 and, hence, was spread over a considerably wide range of the scale with very low means (less than 0.10) and very low standard deviations. The divergence index *(D)*, the measure of the discrepancy between the test takers' and experts' solutions, also had relatively low means and standard deviations. The values ranged from .01 to .71 and covered a moderately wide range of the scale with a tendency toward low values. The values for correspondence with experts´ solutions (*C*) had high means with low standard deviations and covered a moderate range of the scale.

Table 3

*Index descriptives of different texts*

| | | Precision (*P*) | | | | Overplus (*O*) | | | | Correspondence (*C*) | | | | Divergence (*D*) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Text | *n* | *Min* | *Max* | *M* | *SD* | *Min* | *Max* | *M* | *SD* | *Min* | *Max* | *M* | *SD* | *Min* | *Max* | *M* | *SD* |
| 1 | 470 | 0.00 | 1.00 | 0.21 | 0.30 | 0.00 | 0.58 | 0.09 | 0.09 | -0.29 | 0.87 | 0.65 | 0.17 | 0.02 | 0.61 | 0.14 | 0.08 |
| 2 | 467 | 0.00 | 1.00 | 0.31 | 0.28 | 0.00 | 0.73 | 0.09 | 0.08 | -0.42 | 0.94 | 0.76 | 0.17 | 0.03 | 0.71 | 0.12 | 0.08 |
| 3 | 521 | 0.00 | 1.00 | 0.59 | 0.27 | 0.00 | 0.47 | 0.04 | 0.06 | 0.06 | 0.99 | 0.89 | 0.12 | 0.01 | 0.47 | 0.06 | 0.06 |
| 4 | 468 | 0.00 | 1.00 | 0.43 | 0.23 | 0.00 | 0.60 | 0.09 | 0.09 | -0.05 | 0.87 | 0.68 | 0.13 | 0.07 | 0.53 | 0.16 | 0.07 |
| 5 | 526 | 0.00 | 0.92 | 0.47 | 0.29 | 0.00 | 0.52 | 0.05 | 0.06 | 0.03 | 0.98 | 0.80 | 0.11 | 0.01 | 0.49 | 0.10 | 0.06 |
| 6 | 365 | 0.00 | 1.00 | 0.81 | 0.32 | 0.00 | 0.30 | 0.09 | 0.06 | 0.26 | 0.96 | 0.81 | 0.11 | 0.02 | 0.37 | 0.10 | 0.05 |

*Note*. Subjects worked on 3 booklets with 1-2 highlighting tasks each, hence with 4-5 highlighting tasks overall.

Table 4

*Psychometric properties of all major variables*

| Variable | | *n* | *M* | *SD* | Range | | Intercorrelations | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Potential | Actual | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 1 | *P* | 888 | 0.45 | 0.22 | 0-1 | 0.00-1.00 | **.53[a]** | .05 | -.24* | .26* | .05 | .04 | .07* | .25* |
| 2 | *O* | 888 | 0.08 | 0.07 | 0-1 | 0.00-0.57 | --[d] | **.77[a]** | .94* | -.94* | .00 | .00 | -.11* | -.26* |
| 3 | *D* | 888 | 0.11 | 0.06 | 0-1 | 0.02-0.52 | --[d] | --[d] | **.74[a]** | -.99* | -.01 | -.01 | -.12* | -.33* |
| 4 | *C* | 888 | 0.76 | 0.13 | -1-1 | -0.04-.096 | --[d] | --[d] | --[d] | **.74[a]** | .01 | .02 | .12* | .33* |
| 5 | *MS* | 930 | 2.76 | 0.57 | 1-4 | 1.00-4.00 | .17* | -.05 | -.07 | .07 | **.66[b]** | .11* | .02 | .06* |
| 6 | *CS* | 930 | 2.29 | 0.69 | 1-4 | 1.00-4.00 | -.02 | .05 | .05 | -.05 | .15* | **.66[b]** | -.02 | .01 |
| 7 | *MR* | 929 | 3.01 | 0.67 | 1-4 | 1.00-4.00 | .09 | -.15* | -.17* | .17* | .11* | -.06 | **.82[b]** | .11* |
| 8 | *RE* | 886 | 0.05 | 1.13 | -∞-∞ | -3.74-4.90 | .49* | -.39* | -.51* | .51* | .16* | -.06 | .17* | **.98[c]** |

*Note*. *P* = Precision, *O* = Overplus, *C* = Correspondence with experts, *D* = Divergence from experts, *MS* = Metacognitive strategies (regulation), *CS* = Cognitive strategies (elaboration), *MR* = Management of internal resources (concentration), *RE* = reading competence; reliability is shown diagonally in bold: [a]McDonald´s Omega from factor analysis´ results (McNeish, 2018), [b]Cronbachs Alpha, [c]estimated for the whole pool of items; manifest correlations are shown above (Pearson)/ below are latent correlations (reading only as manifest variable); [d]structural equation models (latent correlations) with unsatisfactory fit; one-tailed level of significance * $p < .05$.

Our assumption that the indices represented different scales was investigated by confirmatory factor analyses. Hence, for all indices except for precision, the factors can be assumed to have an acceptable model fit (Overplus: CFI = .99, RMSEA = .01; Correspondence: CFI = .97, RMSEA = .03; Divergence: CFI = .97, RMSEA = .03; see Table A.1 in the Appendix) and reliability. Only the index precision was scaled with unsatisfactory reliability (cf. Table 4). Therefore, this index will be excluded from further analyses in this study. For the indices overplus, correspondence and divergence, the Omega reliabilities (McNeish, 2018) were of a satisfactory level (≥ .74). Nevertheless, the analyses did not reveal one major highlighting factor, with all of these or even selected indices as subfactors.

The descriptive results are shown in Table 4, in particular, the general means, standard deviations and correlations of all major variables. The general means for overplus and divergence were relatively low with respect to the ranges of the scales; the respective standard deviations were also relatively low. However, the general mean for correspondence was comparatively high within the scales range and had a low standard deviation. Hence, the variance of overplus, divergence and correspondence was very low. Probably for this reason, the structural equation models with latent correlations of these indices did not converge. Instead, the manifest correlations were strong and significant, positively directed between overplus and divergence and negatively directed between correspondence and overplus as well as divergence.

### 5.2 Relation of instructed highlighting to reading competence

The correlations between reading competence and highlighting indices were remarkably high, particularly the latent correlations. Here, the weakest significant negative relation was found between reading competence and overplus ($r$ = -.39; $p$ =

.00). The correspondence (*r* = .51; *p* = .00) with and divergence from (*r* = -.51; *p* = .00) the experts´ solutions were significantly related to reading competence to the same extent, but in opposite directions.

### 5.3 Relation of instructed highlighting to the use of strategies

The correlations of the quality indices of instructed highlighting and the use of learning strategies and reading competence are also shown in Table 4. With a focus on the relations of the use of strategies, significant latent correlations of highlighting indices were only found with strategies of concentration. These correlations were low but in the expected direction; in particular, overplus (*r* = -.15; *p* = .00) and divergence (*r* = -.16; *p* = .00) were negatively correlated with concentration, which was positively correlated with correspondence (*r* = .16; *p* = .00). The manifest models showed similar patterns with slightly lower correlations of strategies of concentration with the highlighting indices.

## 6 Discussion

### 6.1 Description of instructed highlighting by means of quality indices

The first objective of this study was to describe highlighting by means of quality indices. Therefore, the results in Figure 1 revealed first insights into the amount of highlighting used in the text units. For the precision index, the values covered a broad range of the scale; the scale was modelled with an acceptable fit and, hence, the scale successfully differentiated between the answers of individuals. Nevertheless, these values have been excluded from analyses because of the unsatisfactory reliability of the assessment of precision. One reason of this unsatisfying reliability might be the fact, that no person of all test takers underlined the solution of text 5 completely. This could be caused by a very high difficulty of the task, ab error in data processing or even in the master solution which cannot be examined with the available data, but gives reason for deeper analyses in future.

In contrast, the overplus index was assessed with a good reliability. The values for this parameter did cover a wide range of the scale, but the means for single texts in general and the standard deviations were close to 0. As shown by these values, the participants mostly highlighted with low overplus and never highlighted the maximum possible additional passages ($O$ = 1). Hence, our findings are inconsistent with the results reported by Peterson (1992) and Leopold and Leutner (2015) and our hypothesis that the values for overplus would vary widely. Furthermore, most subjects highlighted only a few additional passages. One possible reason for these inconsistent results could be due to differences in the instructions. While subjects in this study used highlighting in order to answer a specific and concrete question, individuals in the other studies reported highlighting information that was most important or matched other self-chosen criteria.

The values for the correspondence with experts´ judgements did not cover the full range, but a moderate range of the scale. Therefore, the majority of the participants' solutions tended to correspond with the experts´ judgements to a high degree. In sum, this parameter allowed satisfactory differentiation between the answers of individuals, although the text units´ lengths were not taken into account and, hence, the data might entail some vague information.

By contrast, the text units´ lengths were taken into account through weighting used in the calculation of divergence from experts´ judgements. Therefore, this parameter was expected to provide the most exact measure of the highlighting of the participants. The values of this index were at a low range of the scale with a general tendency toward low or no divergence from experts´ judgements. Nevertheless, this index allowed some differentiation between the test takers´ answers.

Altogether, low variances were found for all indices. This could have been the reason for that a general highlighting factor was not found. Nevertheless, the highlighting behavior of persons working on different texts seemed to be similar to a certain extent, in particular with respect to Overplus as well as Divergence and Correspondence with experts' judgements, as revealed by the factor analyses. However, some small differences between the indices for different texts might also be due to specific features of texts or tasks; that is, the questions could have influenced the test takers´ highlighting. For example, a closer consideration revealed that the texts differed in complexity according to the readability index of Björnsson (1971) and with regard to their cognitive demands (see also Table A.2 in the Appendix). Whereas Text 1 had a high level of complexity, Text 6 was only of low complexity. Furthermore, the cognitive demands for Text 1 were of Type c) "evaluation and interpretation", whereas Text 6 required cognitive operations of Type a) "identification of single facts" (**Gehrer** et al., 2013). Finally, these features of texts and tasks might have caused differences in the values of correspondence with (Text 1: $M_C$ = .65; $SD_C$ = .17; Text 6: $M_C$ = .81; $SD_C$ = .11) and divergence from the experts´ judgements (Text 1: $M_D$ = .14; $SD_D$ = .08; Text 6: $M_D$ = .10; $SD_D$ = .05) and, hence, their correlation with reading competence. Therefore, complexity of texts and cognitive demands of given tasks should be taken into account and controlled in future analyses. Here, due to limited testing time, only a small number of texts and highlighting tasks with different cognitive demands was implemented.

The intercorrelations between the corroborated scales for overplus as well as correspondence with and divergence from experts´ judgements were mostly high and significant and in the expected directions. In detail, overplus was significant negatively correlated with correspondence and significant positively associated with divergence: The less the subjects highlighted additional passages, the more their solution corresponded with and the less it diverged from the experts´ judgements. Furthermore, divergence and correspondence were in an opposed relation: The less divergence was found, the stronger correspondence with experts´ judgements, as expected.

### 6.2 Relation of instructed highlighting to reading competence

Regarding the relation of reading competence with the quality aspects of instructed highlighting, the correlations were remarkably high and in the expected directions: The higher the number of additional words subjects highlighted, the lower their reading competence, which is in accordance with the hypotheses. Furthermore, participants with high reading results highlighted with a low amount of overplus and a high level of concordance with the experts´ solutions and a low level of divergence from them, as expected. Finally, as hypothesized, participants with low reading results highlighted with more overplus and also deviated to a larger extent from ideal highlighting. When comparing the different indices in their relation to reading competence, we found that the correlations of divergence from and correspondence with experts´ judgements were equally high, but in opposite directions. Therefore, both parameters might be applicable as indicators for reading, while the consideration of text units' lengths in index *D* did not seem to provide an additional advantage.

### 6.3 Relation of instructed highlighting to the use of strategies

The expected correlations between instructed highlighting indices and use of strategies occurred rarely. Only strategies of concentration correlated at a low level with highlighting indices, but less than expected. Hence, the more individuals reported using strategies of concentration, the less overplus they showed, the less their highlighting diverged from the experts´ judgements and the more it corresponded with the experts’ judgements, as expected. These results support the assumption that the selection of specific text information during highlighting is performed better if it is accompanied by more activities that facilitate attention and concentration as one type of resource-related strategy. Nevertheless, significant correlations were not found between regulation and elaboration and highlighting. These results were more in line with the hypothesis for elaboration, since it involves processes that go beyond superficial text features, which are in focus of highlighting. With regard to regulation, the results did not support our assumption that we would find strong positive correlations. Therefore, the outcomes did not reinforce the assumption that higher regulation of reading, for example, repeated reading of text passages if the reader lacks understanding (e.g., Friedrich & Mandl, 1997; Pintrich & Garcia, 1994) enhances quality aspects of highlighting. Finally, the outcomes did not support the assumed link between the metacognitive capabilities of individuals and their highlighting with respect to overplus and precision (Leopold & Leutner, 2015).

One reason for these unexpected results and the low correlations in general might be the use of scales with only three items. Because of their shortness, their explanatory power for assessing the use of learning strategies might have been limited. Supporting this assumption, we did not find a positive correlation between learning strategies and reading competence, which had been shown in many studies (e.g., Artelt et al, 2001; Leopold et al., 2006). One further reason for the unexpected results could be a problem of the validity of self-reports on the frequencies of using strategies, as mentioned before. In previous studies (e.g., Winne & Jamieson-Noel, 2002), the explanatory power of self-reports on frequencies of using strategies is criticized increasingly. In fact, Veenman (2005) argued that there is little or no correspondence between selfreported strategies and actual behavior. Hence, subjects who report the use of learning strategies might not use them or even not in appropriate situations. But this (non-)agreement between self-reports and real learning behavior also might differ depending on the type of strategy according to the above-mentioned results of Brennan and colleagues (1986). Nevertheless, reports on the frequencies of using learning strategies do not provide appropriate measurements for the quality of the use of strategies (e.g., **Artelt**, 2000). Instead of self-reports on using strategies, tests of declarative knowledge on strategies seem to offer a more valid measure of strategical competence, and hence, are promising to uncover the relations of capabilities of highlighting and the use of learning strategies.

## 6.4 Outlook

One core contribution of this study is the introduction of indices of quality aspects of instructed highlighting which provide objective, valid and comparable parameters across texts for describing highlighting behavior. These indices have revealed initial indications about the underlying capabilities of highlighting as well as results on its relation to use of strategies and reading competence. The most important finding is that highlighting behavior has a low relation with use of strategies of concentration but correlates strongly with reading competence. Therefore, highlighting results might indicate capabilities of reading which are accompanied by concentration. Furthermore, specific quality aspects of highlighting promise to be remarkable predictors of reading competence, in particular the indices divergence from (*D*) and correspondence with experts´ judgements (*C)*. Therefore, the implementation of highlighting items within the assessment of reading competence might be informative, in particular if the importance of these quality aspects can be verified in further analyses. In order to reinforce this

assumption, a few consequences for further research have to be considered.
One of these consequences for future analyses is that highlighting behavior has to be captured by means of symbol-based coding (instead of text units-based coding). By using methods with a higher level of discrimination, we could achieve higher sensitivity to the interesting criteria and differences of the participants´ answers, and, hence, obtain new insights into the factor structures of highlighting capabilities and their relations to the properties of readers. Based on such data, in particular correspondence with experts´ judgements is expected to provide more detailed results, whereas the advantage of weighting text units' lengths using index *D* will be less important.
A second consequence for further investigation on the relation between instructed highlighting and strategy use is that the assessment of learning strategies should be conducted by means of more sophisticated methods. As reported, the assessment of cognitive and metacognitive learning strategies was conducted via questionnaires (with only three items). Therefore, further analyses should be based on data on the use of strategies with higher reliability, e.g., via observation or testing of declarative knowledge on strategies. Furthermore, the quality (cf. Leopold & Leutner, 2015) as well as the appropriateness of the use of learning strategies has not yet been taken into account, which might also have been a cause for the unexpected results. From a theoretical perspective, strong relations are also expected between instructed highlighting and rehearsal and monitoring strategies, which were not assessed within this study. Hence, for further research, it would be promising to assess learning strategies using additional scales on rehearsal and monitoring.
A third consequence which promises to generate new insights into the relation between reading competence and highlighting indices is controlling for the influences of

properties of texts and tasks. By controlling both in future analyses, we expect to find stronger relations between highlighting indices and reading competence.

## 7 Conclusion

In summary, the results support the assumption that instructed highlighting requires many operations of reading, which are mastered approximately as well as reading operations are performed by readers. Furthermore, the results indicate that "good highlighting" is associated with a certain amount of concentration. Regarding the remaining learning strategies, the expected correlations – and even the assumed heights – were not found. Therefore, our results do not support the assumption that the processes of instructed highlighting are primarily similar to the operations of using specific learning strategies. Instead, highlighting probably should be seen as a *study aid* and therefore as one procedural element of strategy use whereas goal-orientated, metacognitively regulated and conscious strategic learning as defined above requires much more. Finally, instructed highlighting has a strong relation to reading competence and a low correlation with concentration. Therefore, the results provide first indications in favour of our assumption that specific quality aspects of instructed highlighting might be applicable as indicators of capabilities of reading. Hence, highlighting items might offer an innovative, activating and interactive format within technology-based reading assessment. If these competences measured by highlighting represent reading competence of the underlying concept or indicate a more specific competence of *focused reading* which is accompanied by a certain amount of concentration has to be evaluated in further analyses.

**References**

Artelt, C. (2000). *Strategisches Lernen*. Münster, Germany: Waxmann.

Artelt, C. (2006). Lernstrategien in der Schule. In H. Mandl, & H. F. Friedrich (Eds.), *Handbuch Lernstrategien* (pp. 337–351). Göttingen, Germany: Hogrefe.

Artelt, C., Demmrich, A., & Baumert, J. (2001). Selbstreguliertes Lernen. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann, & M. Weiß (Eds.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 271–298). Opladen, Germany: Leske + Budrich.

Björnsson, C.-H. (1971). *Lasbarhet (Readability)*. Kopenhagen, Denmark: Gad. (Original Work published 1968).

Blossfeld, H.-P., & von Maurice, J. (2011). Education as a lifelong Process. *Zeitschrift für Erziehungswissenschaft: Sonderheft 14*, 19-34.

Brennan, S., Winograd, P. N., Bridge, C. A., & Hiebert, E. H. (1986). A comparison of observer reports and self-reports of study practices used by college students. *National Reading Conference Year Book, 35*, 353–358.

Eckert, A. (1998). *Kognition und Wissensdiagnose. Die Entwicklung und empirische Überprüfung des computergestützten wissensdiagnostischen Instrumentariums Netzwerk-Elaborierungs-Technik (NET)*. Lengerich, Germany: Pabst Science Publishers.

Friedrich, H. F., & Mandl, H. (1997). Analyse und Förderung selbstgesteuerten Lernens. In F. E. Weinert, & H. Mandl (Eds.), *Enzyklopädie der Psychologie, Bd. 4, Psychologie der Erwachsenenbildung* (pp. 237–293). Göttingen, Germany: Hogrefe.

Gehrer, K., Zimmermann, S., Artelt, C., & Weinert, S (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online, 5*, 50-79.

Green, D. M., & Swets J. A. (1966). *Signal Detection Theory and Psychophysics*. New York, NY: Wiley.

Heyne, N. (2015). Video-based analyses of reading instruction in the fourth grade based on the observation system LUPE. In W. Schnotz, A. Kauertz, H. Ludwig, A. Müller, & J. Pretsch (Eds.), *Multidisciplinary Research on Teaching and Learni*ng (pp. 167–183). Basingstoke, UK: Palgrave Macmillan.

Holley, J.W., & Guilford, J.P. (1964). A note on the G-index of agreement. *Educational Psychological Measurement. 24*, 749–753.

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

Kobayashi, K. (2009). Comprehension of relations among controversial texts: effects of external strategy use. *Instructional Science, 37,* 311–324.

Lenhard, W., & Artelt, C. (2009). Komponenten des Leseverständnisses. In W. Lenhard & W. Schneider (Eds.), *Diagnose und Förderung des Leseverständnisses: Tests und Trends* (pp. 1–18). Göttingen, Germany: Hogrefe.

Lenhard, W., & Lenhard, A. (2014-2017). *Berechnung des Lesbarkeitsindex LIX nach Björnson*. Bibergau, Germany: Psychometrica. Retrieved from: http://www.psychometrica.de/lix.html.

Leopold, C., & Leutner, D. (2015). Improving students' science text comprehension through metacognitive self-regulation when applying learning strategies. *Metacognition & Learning, 10*, 313–346.

Leopold, C., den Elzen-Rump, V., & Leutner, D. (2006). Selbstreguliertes Lernen aus Sachtexten. In: M. Prenzel & L. Allolio-Näcke (Eds.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms* (pp. 268–288). Münster, Germany: Waxmann.

Mayer, R. E. (1996). Learning strategies for making sense out of expository texts: The SOI model for guiding the cognitive processes in knowledge construction. *Educational Psychology Review, 8,* 357–371.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods,, 23*(3), 412–433. http://dx.doi.org/10.1037/met0000144

Muthén, L. K., & Muthén, B. O. (1999-2012). *Mplus (Version 7)*. Los Angeles, CA: Muthén & Muthén.

Peterson, S. E. (1992). The cognitive functions of underlining as a study technique. *Reading Research and Instruction, 31*, 49–56.

Pintrich, P.R., & Garcia, T. (1994). Self-regulated learning in college students: Knowledge, strategies, and motivation. In P.R. Pintrich, D.R. Brown, & C.E. Weinstein (Eds.), *Student motivation, cognition, and learning* (pp. 113–133). Hillsdale, NJ: Erlbaum.

Pohl, S., & Carstensen, C. H. (2013). Scaling of the competence tests in the National Educational Panel Study – Many questions, some answers, and further challenges. *Journal for Educational Research Online (JERO)*, *5*, 189–216.

Ponce, H. R., & Mayer, R. E. (2014). An eye movement analysis of highlighting and graphic organizer study aids for learning from expository text. *Computers in Human Behaviour, 41*, 21–32.

Sukkarieh, J. Z., von Davier, M., & Yamato, K. (2012). From Biology to Education: Scoring and Clustering Multilingual Text Sequences and Other Sequential Tasks. Research Report. Retrieved from http://dx.doi.org/10.1002/j.2333-8504.2012.tb02307.

Veenman, M. V. J. (2005). The assessment of metacognitive skills: What can be learned from multi-method designs? In C. Artelt, & B. Moschner (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. 77–99). Berlin, Germany: Waxmann.

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., & Carstensen, C. H. (2011). Development of competencies across the life span. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), Education as a lifelong process: The German National Educational Panel Study (NEPS) *Zeitschrift für Erziehungswissenschaft, 14,* pp. 67–86. Wiesbaden, Germany: Springer VS.

Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. C. Wittrock (Ed.), *Handbook of research in teaching* (pp. 315–327). New York, NJ: Macmillan.

Weinstein, C. E., Husman, J., & Dierking, D. R. (2000). Self-regulation interventions with a focus on learning strategies. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 727–747). San Diego, CA: Academic Press.

Wild, K.-P. (2006). Lernstrategien und Lernstile. In D. H. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* (3rd ed., pp. 427–432). Weinheim, Germany: Beltz Psychologie Verlags Union.

Wild, K.-P., & Schiefele, U. (1994). Lernstrategien im Studium: Ergebnisse zur

Faktorenstruktur und Reliabilität eines neuen Fragebogens. *Zeitschrift für Differentielle und Diagnostische Psychologie 15*, pp. 185–200.

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students´ calibration of self-reports about study tactics and achievement. *Contemporary Educational Psychology, 27*, 551–572.

# Appendix

Table A.1

*Results of confirmatory factor analyses on highlighting indexes*

| Scale | $\chi^2$ | df | CFI | TLI | RMSEA | SRMR |
|---|---|---|---|---|---|---|
| Precision | 11.89 | 9 | .96 | .93 | .02 | .04 |
| Overplus | 10.41 | 9 | .99 | .99 | .01 | .04 |
| Correspondence | 14.56 | 9 | .97 | .94 | .03 | .05 |
| Divergence | 14.72 | 9 | .97 | .94 | .03 | .05 |

*Note.* Fit-Indizes of 1-factor-models; for $\chi^2$ *of Model Fit* * < $p$ = 0.5 (no case); CFI = Comparative-Fit-Index; TLI = Tucker-Lewis-Index; RMSEA = Root Mean Square Error Of Approximation; SRMR = Standardized Root Mean Square Residual.

Table A.2

*Features of texts and highlighting tasks*

| Text | Length | Solution length | Solution amount | Average sentence length[a] | Amount of long words[a] | Complexity (LIX) [a] | Cognitive demands |
|---|---|---|---|---|---|---|---|
| 1 | 1402 | 97 | 0.07 | 21.70 | 40.20 | 61.90 | 3 |
| 2 | 1752 | 88 | 0.05 | 25.00 | 32.40 | 57.40 | 2 |
| 3 | 1713 | 68 | 0.04 | 13.00 | 40.10 | 53.10 | 1 |
| 4 | 1447 | 199 | 0.14 | 21.30 | 16.70 | 38.10 | 3 |
| 5 | 1482 | 142 | 0.10 | 16.10 | 31.80 | 48.00 | 2 |
| 6 | 1588 | 166 | 0.11 | 15.70 | 30.50 | 46.20 | 1 |

*Note.* Length (*T*) and Solution length (*R*) based on number of signs; Solution amount (*T*/*R*) with range 0-1; Average sentence length in words; Amount of long words (> 6 letters) in percent; Complexity/ Readibility index (LIX; Björnsson, 1968); Cognitive demands classification (Author 4 et al., 2013); [a]calculated by psychometrica (Lenhard & Lenhard, 2014-2017).

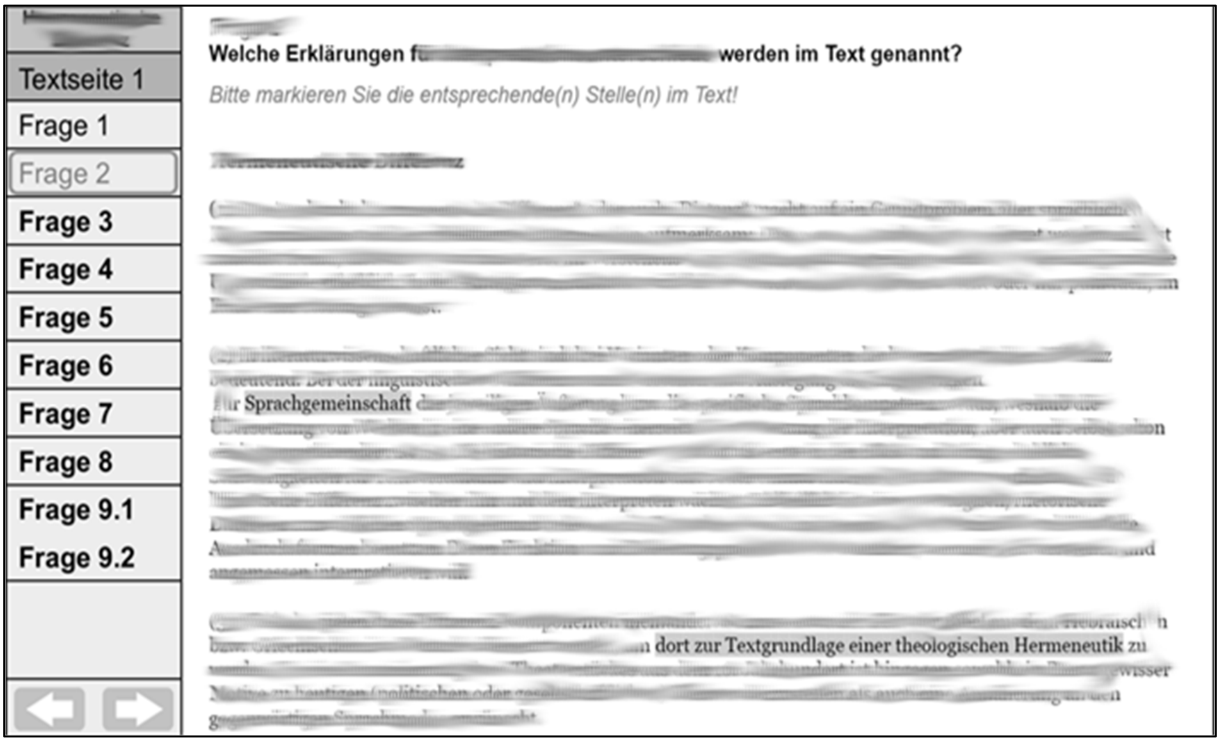


*Figure A.1.* Task with instructed highlighting – example from a development study (Question: “Which explanations are mentioned in the text? Please mark the corresponding passage(s) in the text!”)

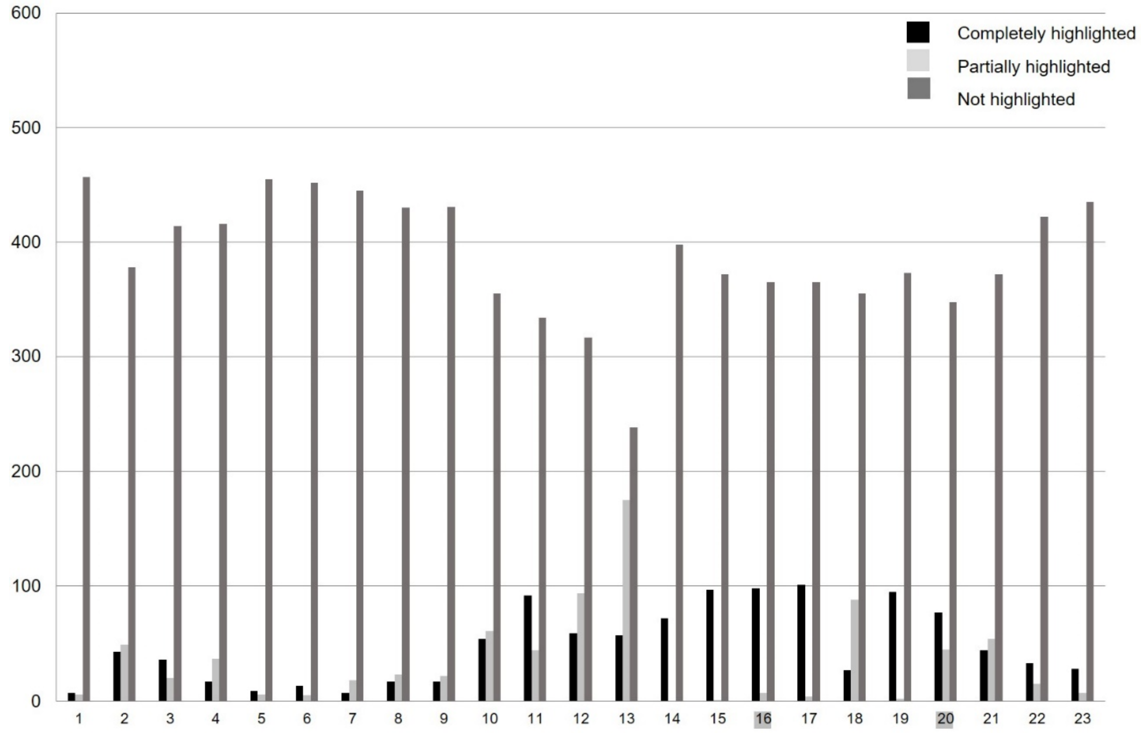


*Figure A.2.* Instructed highlighting of subjects in Text 1 (correct units by experts´ solutions indicated by grey boxes; $N_{\text{Missings}}$ = 467)

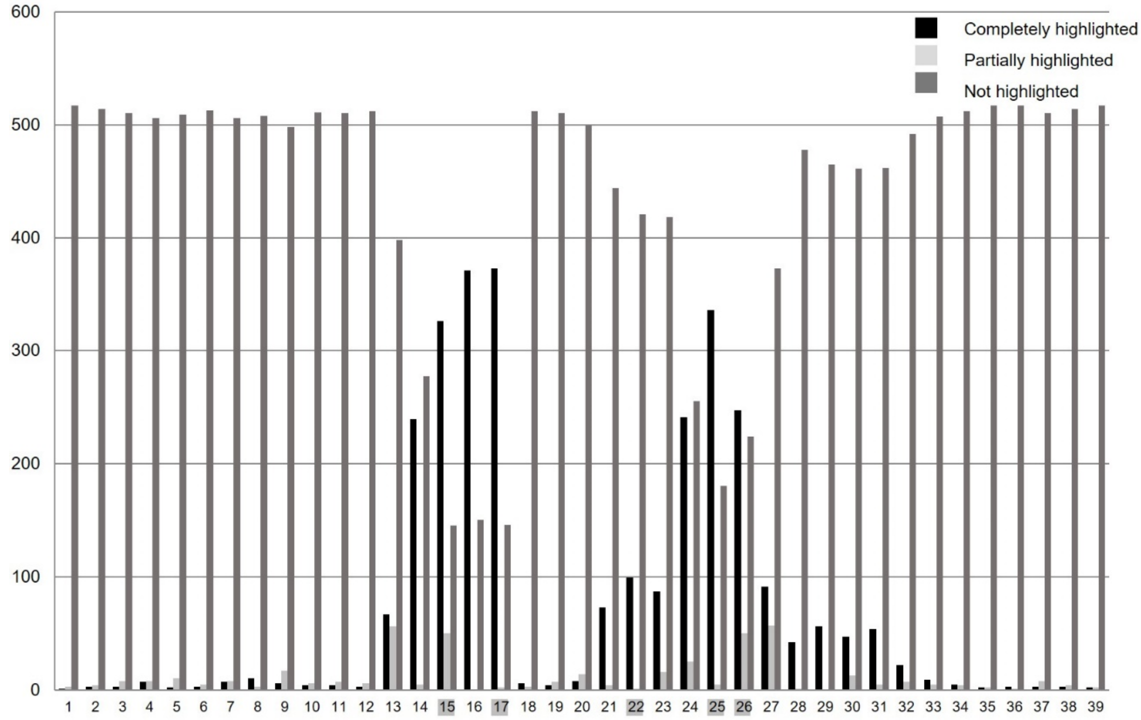


*Figure A.3.* Instructed highlighting of subjects in Text 3 (correct units by experts´ solutions indicated by grey boxes; $N_{\text{Missings}}$ = 416)

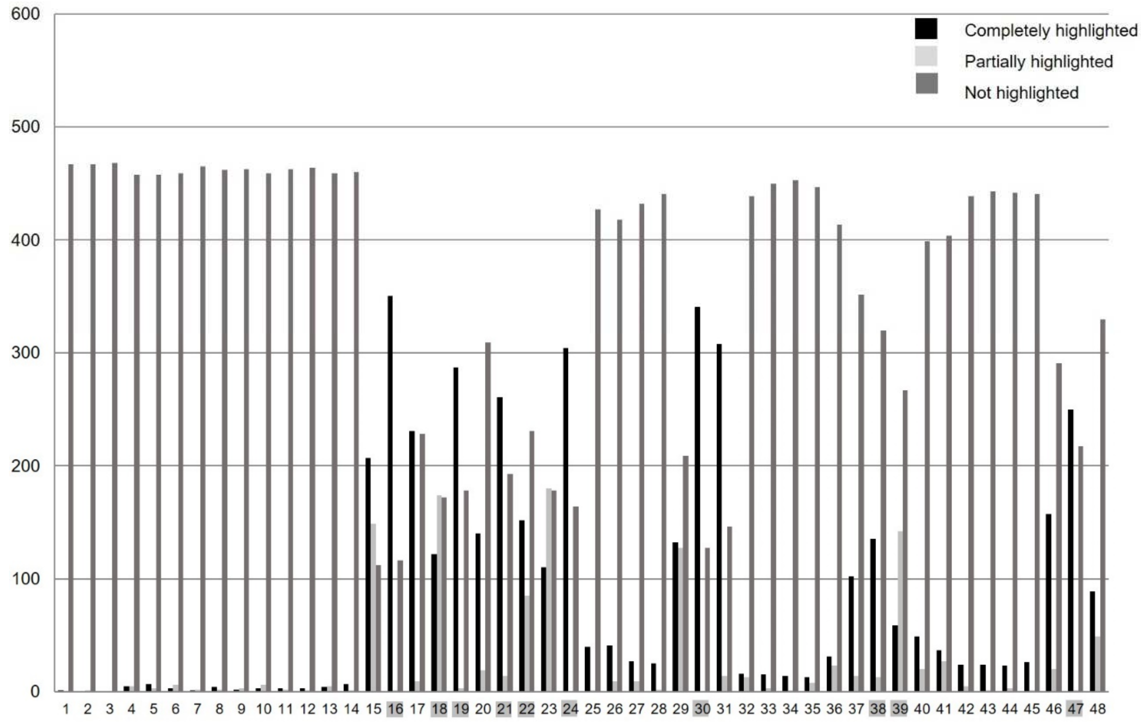


*Figure A.4.* Instructed highlighting of subjects in Text 4 (correct units by experts´ solutions indicated by grey boxes; $N_{\text{Missings}} = 469$)

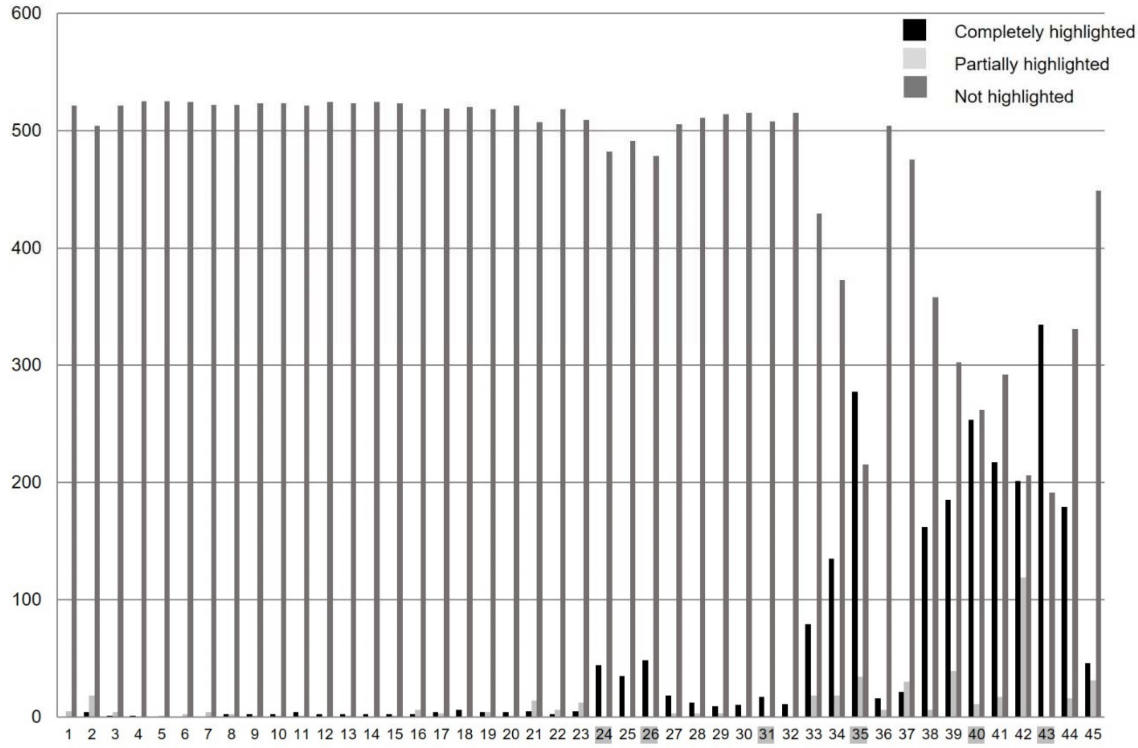


*Figure A.5.* Instructed highlighting of subjects in Text 5 (correct units by experts´ solutions indicated by grey boxes; $N_{\text{Missings}}$ = 411)

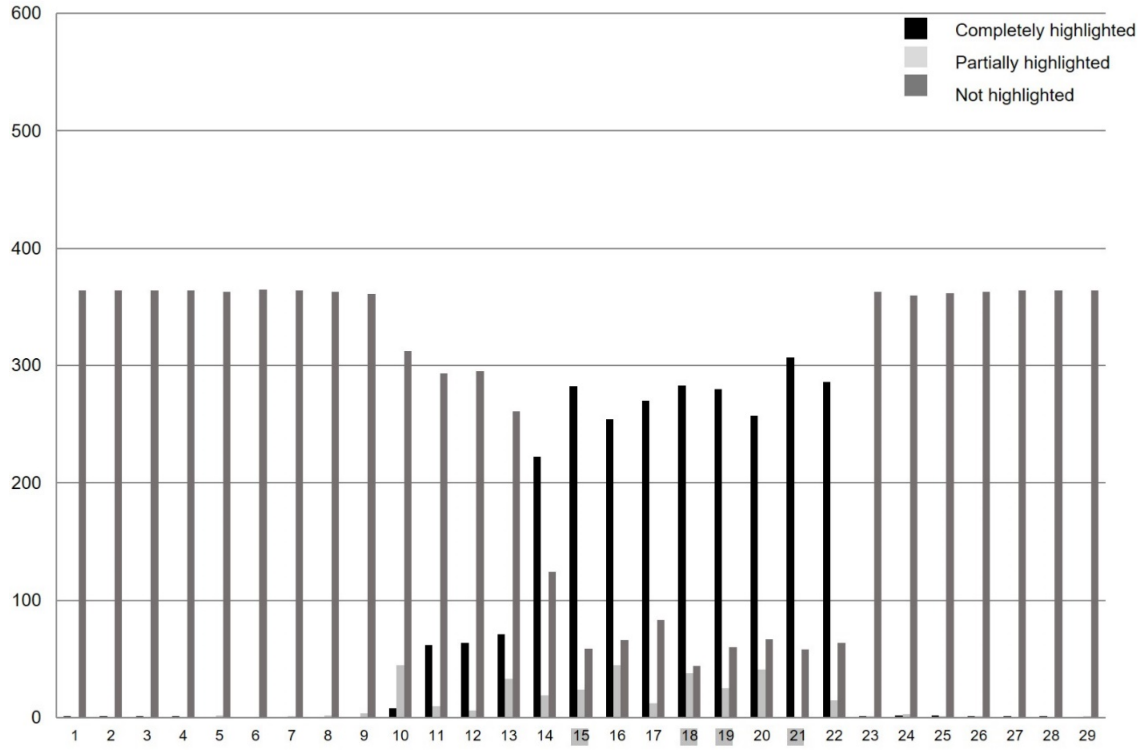


*Figure A.6.* Instructed highlighting of subjects in Text 6 (correct units by experts´ solutions indicated by grey boxes; $N_{\text{Missings}}$ = 572)