Running Head: PERCEIVED TASK DEMANDS IN MULTIPLE DOCUMENTS

Readers' Perceived Task Demands and their Relation to Multiple Document Comprehension Strategies and Outcome

Cornelia Schoor[a], Jean-François Rouet[b], Cordula Artelt[c, f], Nina Mahlow[c], Carolin Hahnel[d, e], Ulf Kröhne[d] & Frank Goldhammer[d, e]

[a] University of Bamberg, Department of Educational Research, Markusplatz 3, 96047 Bamberg, Germany

[b] Center for Research on Cognition and Learning, MSHS - Bâtiment A5, 5, rue T. Lefebvre - TSA 21103, 86073 Poitiers cedex 9, France

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

[d] DIPF | Leibniz Institute for Research and Information in Education, Rostocker Straße 6, 60323 Frankfurt am Main, Germany

[e] Centre for International Student Assessment (ZIB), Frankfurt am Main, Germany

[f] University of Bamberg, Department of Longitudinal Educational Research, Wilhelmsplatz 3, 96047 Bamberg, Germany

Post print version of the article: Schoor, C., Rouet, J.-F., Artelt, C., Mahlow, N., Hahnel, C., Kroehne, U., & Goldhammer, F. (2021). Readers' perceived task demands and their relation to multiple document comprehension strategies and outcome. *Learning and Individual Differences, 88*, 102018. https://doi.org/10.1016/j.lindif.2021.102018

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit https://creativecommons.org/licenses/by-nc-nd/4.0 or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Acknowledgements

The project on which this report is based was funded by the German Federal Ministry of Education and Research, funding code 01PK15008. Responsibility for the contents of this publication rests with the authors. In addition, we want to thank the Center for Bavarian-French university cooperation (Bayrisch-Französisches Hochschulzentrum) for granting a mobility allowance to support this paper.

Abstract

Recent research suggests that readers' subjective task understanding influences reading processes and outcomes. Therefore, the present study's aim was to investigate whether the task demands that readers retrospectively report relate to multiple document comprehension strategies and outcome. A total of 310 university students completed three units from a standardized multiple-document comprehension test and answered an open-ended task demands question after each unit. Amongst others, participants comprehended single- and multiple-document activities to be task demands. Comprehending deep-level single-document activities and management activities to be task demands related to corroboration and proactive sourcing, respectively. However, comprehending multiple-document activities to be task demands was related neither to students' multiple-document comprehension nor to their realized multiple-document activities. The data suggest a context schema formation across test units: In later units the participants comprehended more often multiple-document activities and less often surface-level single-document and management activities to be task demands, and conducted more sourcing.

Key words: multiple document comprehension; task model; reading; strategies; log-file data

Readers' Perceived Task Demands and their Relation to Multiple Document Comprehension Strategies and Outcome

1 Introduction

Every teacher probably knows that students do not always understand task instructions as they were intended. Inter-individual differences in one's understanding of a task can be interpreted as differences in the mental representation of task demands. This so-called *task model* is supposed to develop during the work on the task (Rouet et al., 2017). However, little is known about how students understand tasks that require multiple document comprehension (MDC). MDC is necessary, for example, when students prepare a seminar talk or write their thesis. Beyond the understanding of single texts, such tasks require the comparison and integration of information across texts and a representation of "who said what" (Perfetti et al., 1999; Wiley et al., 2014). Although research investigating task effects on comprehension provides indirect evidence for the existence of task models, research on the task's representation is scarce (List et al., 2019).

The present study aims to investigate the task demands that university students report after completing a multiple document assignment by relating these demands to the students' behavior during work and to their MDC. It is based on the underlying assumptions that 1) students differ in their task understanding, and 2) that an appropriate understanding of MDC tasks and their demands includes the belief that multiple-document activities, such as comparing information across texts (Cerdán & Vidal-Abarca, 2008; Wineburg, 1991) or considering sources (Kammerer et al., 2016; Mason et al., 2014), are necessary. In the following, we use the term "perceived task demands" in order to refer to the activities that readers understand and belief to be necessary to successfully complete the task.

1.1 The Task Model in Task-Oriented Reading

Reading is a purposeful activity (Britt et al., 2018; Snow & the RAND Reading Study Group, 2002). Reading purposes might be shaped by internally set goals (e.g., reading for entertainment or to inform oneself about a topic to make a personal decision or to solve a problem) or by an externally set task (e.g., an academic assignment) (see Lorch et al., 1993). The present research focuses on the latter. Readers' purposes influence reading processes and outcomes, both in the comprehension of single (e.g., McCrudden & Schraw, 2007; Rouet, 2006; Van den Broek et al., 2001) and multiple texts (e.g., Bråten & Strømsø, 2010; Le Bigot & Rouet, 2007; Stadtler et al., 2014; Wiley & Voss, 1999). For example, task instructions requesting the writing of an argumentation or summary lead to better MDC performance compared to instructions for global understanding or keyword listing (Bråten & Strømsø, 2010; Le Bigot & Rouet, 2007; Stadtler et al., 2014). However, the results are ambiguous with regard to whether an argumentation task is more beneficial than a summary task (Hagen et al., 2014; Wiley & Voss, 1999) and there are indications that only readers with high prior knowledge benefit from argumentation tasks (Gil et al., 2010).

Research on task effects has often assumed that students' understanding of task instructions is straightforward. However, task instructions themselves are subject to comprehension processes (e.g., Rouet, 2006), and readers may differ in their ability to understand what is asked from them. According to the RESOLV theory of reading as problem solving (Britt et al., 2018; Rouet et al., 2017), readers interpret a task to build a *task model*, that is, their personal interpretation of the task demands in the form of a mental model (Rouet & Britt, 2011). The task model is influenced by the physical and social context, which is represented in the *context model* (also a mental model) based on features of the request, the requester, the audience, and the self (Britt et al., 2018; Rouet et al., 2017). Accordingly, recent research suggests that the authority of the requester influences the representation of strategies in the task model (Rouet et al., 2020). The context model is assumed to be affected by pre-existing *context schemata* that reflect previously learned experiences about a context (Britt et al., 2018; Rouet et al., 2017; see Lorch et al., 1993, for typical reading context schemata).

The task model includes the mental representation of the means (i.e., activities and strategies) available to achieve an expected outcome (Rouet et al., 2017). These means are based on strategy knowledge and the context model (Rouet et al., 2017). The task model and the activities represented are supposed to influence reading behavior and outcome, as supported by all above-mentioned studies on task effects, which indirectly speak for the impact of the reader's task model on reading. With regard to research on the task model itself, a few studies could already demonstrate its influence on single-text comprehension (Cerdán et al., 2013; Cerdán et al., 2019; Llorens & Cerdán, 2012), showing that the probability to correctly solve a task is higher when students reflect upon their task understanding before beginning the task. For secondary-school students, self-explaining the task seems only beneficial for good, but not for poor readers (Cerdán et al., 2013) while an aid to understand the task helps only poor, but not good readers (Cerdán et al., 2019). These findings indicate that developing an appropriate task model is beneficial for solving text comprehension tasks.

Forming an accurate representation of the task demands may play an even more critical role for more complex reading assignments, such as those that require students to read and integrate information from multiple documents. However, research on multiple document task models is rather scarce. List et al. (2019) analyzed students' perceptions of widespread multiple document tasks and found relationships between task perception and performance in an argument task. However, they did not report whether students represent specific multiple-document activities, such as corroborating information across documents. Students' actual reading strategies (and outcomes) may depend on whether they believe that the task requires them to perform specific activities (Britt et al., 2018).

A significant obstacle when investigating students' task models is how to assess them without disrupting the actual reading process. Since research on the task model is scarce, there is no standard measurement approach so far. Prior research has used both multiple choice questions (Llorens & Cerdán, 2012) and an open format (Cerdán et al., 2013) to assess students' understanding of the demands of comprehension questions. However, questioning students as or before they read may influence their approach to the task. In fact, Llorens and Cerdán (2012) showed that the time of identifying the correct rephrasing of the task instructions out of a list of four alternatives influenced students' task performance: Identifying it before the task lead to a better task performance in comparison to identifying it after the task. This could be an effect of self-explanation (e.g., McNamara, 2017) such that students who self-explain the task instruction before actually working on the task have a more appropriate task model at the beginning of the work and therefore perform better, while those identifying the correct task understanding only after working on the task cannot change their behavior anymore based on their gained understanding. This is also in line with the RESOLV theory according to which students update their task model during engaging in the task. Expanding previous theories which have a rather static notion of task understanding (e.g., Brand-Gruwel et al., 2009; Flavell, 1979; Rouet, 2006; Winne & Hadwin, 1998), RESOLV suggests that the reader's task model might not be complete before actually engaging in the task (Rouet et al., 2017). For both reasons (i.e., influence on the work on the task, potential incompleteness of the task model), assessing the task model before the actual work on the task is not optimal. In the absence of a non-invasive method to monitor the task model as students read, a possibility, although not ideal, is to ask students to self-report their task model just after task completion.

In addition to learning about the task while working on it, RESOLV also suggests that across multiple similar tasks within the same context, learning about the task and the context can occur and result in a context schema (see Winne & Hadwin, 1998). The context schema influences the task model for the next task in the same or similar context (Britt et al., 2018; Rouet et al., 2017). Therefore, the formation or update of a context schema could imply behavioral changes and changes in the task model that occur across several similar tasks (see Lorch et al., 1993). Consequently, readers should report more task-relevant activities as they go through several similar tasks.

1.2 Strategies of Multiple Document Comprehension (MDC)

Beyond single-text comprehension, MDC requires readers to compare and integrate information across documents (Stadtler et al., 2013). The Documents Model Framework (e.g., Britt & Rouet, 2012) assumes that, in addition to cognitive representations of single texts (e.g., Kintsch, 1998), a *documents model* is built that represents an integration of the content and meta-information of the document sources.

Strategies of both single-document (SD) and multiple-document (MD) comprehension positively relate to MDC (e.g., Anmarkrud et al., 2014; Wineburg, 1991; Wolfe & Goldman, 2005). Afflerbach and Cho (2009) distinguish strategies for identifying and learning important information, monitoring, and evaluating. In other approaches, SD strategies have often been divided into *surface-level SD strategies* and *deep-level SD strategies* (e.g., Marton & Säljö, 1976; Murphy & Alexander, 2002). Deep-level SD strategies include a personalization or transformation of text (e.g., linking information to prior knowledge) whereas surface-level SD strategies stick closely to the text surface (e.g., re-reading the text; Murphy & Alexander, 2002). Additionally, *strategies to manage* oneself and/or one's environment (e.g., time management, avoidance of distractions) and *metacognitive strategies* have been proposed (e.g., Pintrich et al., 1993; Weinstein & Mayer, 1986).

In addition to SD strategies, MDC requires specific activities that address the comparison and integration of different documents, such as *sourcing* and *corroboration* (Rouet et al., 1997; Wineburg, 1991). Sourcing refers to the attention to source information of the document (e.g., Bråten et al., 2018) and can serve several functions (Hahnel, Kroehne, et al., 2019). One of them, *proactive sourcing*, is to provide a framework which guides the encoding of text, which can be observed in experts who look at the source information very early in the process of document reading (Wineburg, 1991). In contrast, *repeated sourcing* results from the need to update memory traces of source information (Hahnel, Kroehne, et al., 2019). Accordingly, it has been shown that the number of conflicts across documents that are supposed to make sourcing more likely (Braasch et al., 2012; Stadtler & Bromme, 2014) is positively related to the behavior of re-accessing source information (Hahnel, Kroehne, et al., 2019).

Corroboration refers to the comparison of content across documents (Wineburg, 1991). It seems to be related to the belief that documents do not convey "truth" but rather arguments, and that seeming facts must be corroborated by another source (Hynd et al., 2004). However, corroboration has not received the same attention by research as sourcing yet.

Research with readers of several ages indicates that those who are more proficient in MDC also use more strategies of both SD and MD comprehension (e.g., Anmarkrud et al., 2014; Wineburg, 1991; Wolfe & Goldman, 2005). Even for university students, dealing with multiple documents is a challenging task (Rouet et al., 1997). In line with this finding, also university students show interindividual differences in the aforementioned strategies of corroboration and sourcing, which might account for differences in their MDC (e.g., Anmarkrud et al., 2014).

1.3 The Present Study

A competent reader of multiple documents can be expected to not only *conduct* MD activities while reading, but also to *construct a task model* in which these activities are represented as task demands, at least after completing work on an MD task (Rouet et al., 2017; see theoretical cases by Britt et al., 2018). Based on the RESOLV theory, we focused on the following open questions:

1. How do activities that readers report as perceived task demands in a retrospective assessment at the end of an MD assignment relate to behavior during the task and its outcome?
2. Is there evidence for a context schema development such that after working on several similar MD assignments, readers more often perceive MD activities as task demands in a retrospective assessment?
3. Is there indirect evidence for a context schema development such that after working on several similar MD assignments, MD activities are conducted more often?

According to these questions and based on RESOLV, we hypothesized:

H1: Readers who perceive MD activities as task demands in a retrospective assessment realize MD activities more frequently in observable behavior (i.e., corroboration and sourcing; *reading behavior hypothesis*).

H2: Readers who perceive MD activities as task demands in a retrospective assessment show a better MDC (*MDC hypothesis*).

H3: The activities perceived as task demands in a retrospective assessment change across assignments such that the perception of MD activities as task demands is more frequent at the end of later than earlier assignments (*change in task model hypothesis*).

H4: The realization of MD activities in observable behavior (i.e., corroboration and sourcing) changes across assignments such that it is more frequent in later than earlier assignments (*change in reading behavior hypothesis*).

2 Method

2.1 Sample and Design

The participants were 310 university students (79.4% female; age: 18 to 34 years, *M* = 21.4, *SD* = 2.72) from two German universities enrolled in different programs within the social sciences and the humanities (mainly first Bachelor's or Master's semester). Participation was voluntary, outside courses, and compensated with 20 € and a lottery ticket for a digital pad. By means of a balanced incomplete block design, each participant worked on three random MDC assignments ("units") out of six from an MDC test. Their order was rotated.

2.2 Procedure

Figure 1 provides an overview of the present study, which was part of a larger project of developing a computer-based MDC test·[1]. No human subjects approval was required by the local institutional review board according to their guidelines, since participants were neither vulnerable groups, nor there was a deception of the participants or physical, psychological,

[1] Therefore, overlapping data has been used for other contributions. Hahnel, Kroehne, et al. (2019) used partially overlapping data from the same sample, since there the indicators of sourcing were analyzed and validated. The present study uses these indicators and relates them to the perceived task demands. Schoor, Hahnel, Artelt, et al. (2020) used the same sample to analyze the MDC test, which in the present study was used as a final test score that was related to the perceived task demands. Mahlow et al. (2020), in contrast, use a different sample and therefore different data. Schoor, Hahnel, Mahlow, et al. (2020) is an overview chapter with no overlapping data. Rouet et al. (2020) also is on different data, that is there is no overlapping data.

or other risks for the participants. The study was conducted in accordance with APA human subjects principles. After an oral introduction and giving written informed consent (following APA human subjects principles), the participants worked only on the computer. First, demographic and other variables were assessed. After an introduction to the functions of the MDC test (the test tutorial), the participants worked on three MDC units. Since asking to report the task model before working on the task might act as an intervention influencing the actual task performance (thereby also endangering the main aim of developing a test), the perception of task demands was assessed retrospectively after each unit by means of an open-ended question. After having completed all three units, the participants were provided with their reimbursement after about two hours.

2.3 Material and Instruments

2.3.1 Test of Multiple Document Comprehension (MDC)

The computer-based MDC test was developed by Schoor and colleagues (Schoor, Hahnel, Artelt, et al., 2020; Schoor, Hahnel, Mahlow, et al., 2020). It comprises six units on different topics from different domains (see Table 1) designed to represent the variety of requirements of multiple documents across domains (see Goldman et al., 2016). The texts present different perspectives on the same topic without major, but with some minor contradictions on detail level. Source information is available (e.g., publication outlet). To avoid effects of prior knowledge, all topics are fictitious (except for the unit “Universe”). The participants were not informed about this, yet it was obvious for one unit (“2134”). The test was implemented with the CBA ItemBuilder (Rölke, 2012). To provide a more authentic reading situation and obtain process data, participants are able to highlight text and comment on the text margin (Figure 2). These functions were used by the participants in the present study, but rarely for multiple-document activities. All test functionalities (i.e., navigation to texts and items, highlighting, commenting, access of source information, leaving the unit) are explained in a video-based tutorial.

Each unit started with an overview on the number of texts and items and provided the readers with a general reading goal (see last column of Table 1). All these reading goals were designed as a concrete but overarching question for an integrative summary of all texts, based on the summary goals used in prior research (e.g., Bråten & Strømsø, 2010; Le Bigot & Rouet, 2007; Stadtler et al., 2014). Due to time restrictions, the participants were asked in only two units ("Universe", "Nothing") to actually write this summary as the first task. After submitting their essay, they were able to access further closed-ended items. The essay was not used for the MDC test score, but for validation purposes. Each unit comprised 11-17 closed-ended items (Table 1). They aimed to assess the main components identified in the MDC literature (e.g., Britt & Rouet, 2012; Wineburg, 1991):

1. corroboration of information across texts,
2. integration of information across texts,
3. comparison of sources and source evaluations across texts,
4. comparison of source-content links across texts.

Constructed in a way that they could not be solved correctly with information from only one text, the items were presented in single-choice formats (i.e., true/false or select one out of four). The texts were available during item processing. Table 2 provides a detailed description of the item requirements and sample items.

The closed-ended items were scored dichotomously (1=correct, 0=wrong). Based on a Rasch model, weighted likelihood estimates (WLE, Warm, 1989) were used to represent MDC ability. Due to misfit and differential item functioning, a number of items had to be excluded and their scores were not used for the estimation of MDC ability. In total, 67 items in five units were used (WLE reliability of .67). In terms of the validity of score interpretation, the MDC test score was shown to significantly correlate with a better GPA (German Abitur, lower numbers indicating better grades: $r = -.44$, $p < .001$) and Master's students received better MDC scores than Bachelor's students (β = .22, p < .001; Schoor, Hahnel, Artelt, et al., 2020; Schoor, Hahnel, Mahlow, et al., 2020).

2.3.2 Perceived Task Demands

The perceived task demands were retrospectively assessed with an open-ended cued-recall question (Llorens & Cerdán, 2012), repeating the unit reading goal (see Table 1) followed by: "Please explain in your own words what this task asked you for. Please also explain what one has to do in order to solve such a task correctly" (originally in German). This question was intended to elicit the activities the participants considered necessary to solve the task (i.e., the perceived task demands). For the present study, these activities were expected to reflect multiple-document (MD) activities since the readers were requested to create an integrative summary at the onset of each unit (i.e., the reading goal). Although the concrete instruction was unit-specific, all of these instructions required on a more abstract level the corroboration and integration of information across texts and the evaluation and comparison of sources and of source-content links across texts in a similar way. Since each of the 310 participants worked on three units, 930 answers were collected.

Based on theoretical considerations with regard to reading strategies (e.g., Marton & Säljö, 1976; Murphy & Alexander, 2002; Pintrich et al., 1993; Weinstein & Mayer, 1986) and MD activities (e.g., Rouet et al., 1997; Wineburg, 1991), four categories of activities mentioned in the answers were distinguished: surface-level single-document (SD) activities, deep-level SD activities, MD activities, and management activities (Table 3). Each answer was coded independently as to whether it contained activities of a category or not (examples in Table 4). Ten answers were used as training cases, 90 answers were coded independently by a second rater. The inter-rater reliability was almost perfect (Cohen's kappa > .79; Table 3). Thirteen answers (1.4%) were excluded because the participants did not provide an answer at all or referred to the answer they gave to a previous assessment (e.g., "about the same as in the last unit").

2.3.3 Indicators of Realized Multiple-Document Activities: Corroboration and Sourcing

For realized MD activities, we chose corroboration and sourcing since they are specific for MDC and they are comparably straightforward to operationalize based on trace data. The following indicators were extracted from the process data collected in log files using the R package *LogFSM* (Kroehne & Goldhammer, 2018; see also Hahnel et al., 2019):

1) Corroboration: Number of switches between texts.

2) Proactive sourcing: Access of source information considered when it happened within the first 10% of the document processing time.

3) Repeated sourcing: Access of source information when the same source information was accessed at least twice.

To create unit-level variables of proactive and repeated sourcing, the percentages of source access per unit and per person were calculated (i.e., a person accessing one out of three sources of a unit repeatedly would have a score of 1/3 for repeated sourcing in this unit). The intercorrelations of these indicators are displayed in Table 5. On the person level, they correlated significantly with MDC (corroboration: r=.30, p<.001; proactive sourcing: r=.23, p<.001; repeated sourcing: r=.32, p<.001), suggesting the reflection of MD activities.

2.4 Data Analysis

Taking into account the nestedness of data (i.e., three MDC units and task model assessments for each person), multi-level modelling was applied with level 1 being the unit and level 2 the person. The correlations of corroboration and sourcing with each other and the MDC test score were estimated by means of a multi-level model with corroboration and sourcing being variables both at the unit level (within-level) and the person level (between-level). Since the MDC test score was estimated as a person ability parameter based on the responses to all units, it is a person-level variable only (between-level).

For testing the hypotheses, a different multi-level model (Models 1-4, see Figure 3) was specified in Mplus 8.4 (TYPE = TWOLEVEL option) for each hypothesis. In all models, the units were included as a control variable (i.e., several unit dummy variables) to account for the variability of texts and topics (Francis et al., 2018) and the MLR estimator was used. Observed variables for realized activities and MDC were grand-mean-centered. The four perceived-task-demand variables were included as categorical variables. In models with these variables (i.e., Models 1-3), the Montecarlo integration method was used. Model 1 (testing the reading behavior hypothesis) was specified as a random-intercept model with random intercepts for the variables of perceived task demands and realized activities. On the unit level, perceived task demands were modelled to be predicted by realized activities. Realized activities were allowed to correlate. Model 2 (testing the MDC hypothesis) was specified as a random-intercept model with perceived-task-demand variables on the unit level being predicted by the unit (control variable). In order to test the hypothesis, the random intercepts on the person level for these variables were correlated with MDC. Model 3 (testing the change in task model hypothesis) was also specified as a random-intercept model with perceived-task-demand variables on the unit level being predicted by unit and position. Position was dummy-coded with the reference category "second position" (i.e., the second out of three units). Thus, the two dummy variables coded the difference between second and first unit and between second and third unit, respectively. Note that due to this type of coding, an increase from position 1 to position 2 would result in a negative coefficient for the first dummy variable, but an increase from position 2 to position 3 would be mirrored by a positive coefficient for the second dummy variable. Model 4 (testing the change in reading behavior hypothesis) was also a random-intercept model with realized activities on the unit level being predicted by unit and position (dummy-coded).

3 Results

A descriptive screening showed that overall multiple-document (MD) activities were mentioned most frequently as task demands (in almost 60% of the answers; Table 3). About 85% of all participants mentioned MD activities at least once. Surface-level single-document (SD) activities were mentioned in about half of the answers; deep-level SD and management activities each in about a quarter of the answers.

3.1 Relationship of Perceived Task Demands with Indicators of Corroboration and Sourcing

The predicted positive relationship of MD activities perceived as task demands with realized activities of corroboration and sourcing (H1, reading behavior hypothesis) was analyzed in Model 1 (main results in Table 6). There were within-person level effects of the unit with regard to activities perceived as task demands and realized activities (βs ranging from -.26 to .18, reference unit “2134”) except proactive sourcing, for which no significant unit effects were found. The perception of deep-level SD activities as task demands was significantly predicted by corroboration (odds ratio [OR] = 1.02, *p* = .037), and the perception of management activities as task demands was significantly predicted by proactive sourcing (OR = 0.46, *p* < .001). No significant relationships were found between the perception of MD activities as task demands and indicators of realized MD activities.

3.2 Relationship of Perceived Task Demands with Multiple Document Comprehension

The relationship of the perception of MD activities as task demands and MDC (H2, MDC hypothesis) was analyzed in Model 2. The results (Table 7) show no significant relationship of perception of MD activities as task demands and MDC (β = .08, *p* = .322). The substantial relationship of MDC with the perception of deep-level SD activities as task demands also fails to reach significance (β = .17, *p* = .066).

3.3 Perceived Task Demands across Assignments

The predicted more frequent perception of MD activities as task demands in later units (H3, change in task model hypothesis) was analyzed in Model 3 (main results in Table 8). As predicted, the perception of MD activities as task demands was more frequent in later units. Precisely, it was less frequent in the first unit compared to the second (OR = 0.63, *p* = .002), whereas there was no significant difference between the second and the third unit (OR = .94, *p* = .751). Moreover, the perception of surface-level SD (OR = 2.28, *p* = .003) and management activities (OR = 2.96, *p* = .006) as task demands was more frequent in the first as compared to the second unit. These findings are illustrated in Figure 4.

3.4 Realized Behavior across Assignments

The realization of MD activities between units (H4, change in reading behavior hypothesis) was analyzed in Model 4 (main results in Tables 9). There were effects of the unit position on proactive sourcing across all three units (position 2 – position1: β = -.23, *p* < .001; position 2 – position3: β = .07, *p* = .032) and on repeated sourcing for the change between first and second unit (position 2 – position1: β = -.11, *p* = .002; position 2 – position3: β = .02, *p* = .507). There was no effect on corroboration (position 2 – position1: β = -.01, *p* = .690; position 2 – position3: β = -.06, *p* = .052). Proactive and repeated sourcing were more frequent in later units, which corresponds to a previous finding obtained with the same sample (Hahnel, Kroehne, et al., 2019).

4 Discussion

The present study examined the activities that university students perceived as task demands after they completed three different multiple-document (MD) assignments and the relationships of these perceived task demands with reading processes and reading outcomes. As expected, the students perceived MD activities, such as comparing texts and sourcing, as task demands. Unexpectedly, the perception of MD activities as task demands was not related to behavioral indicators of MD activities. However, the perception of deep-level single-document (SD) activities as task demands was positively associated with corroboration, and proactive sourcing negatively predicted the perception of management activities as task demands. Moreover, there was no relationship of multiple-document comprehension (MDC) with MD activities perceived as task demands. As expected, repeated exposure to similar MD assignments, such as the MDC units, lead to an increased perception of MD activities as task demands and an increase in sourcing behavior.

4.1 Relationship of Activities Perceived as Task Demands with Realized Activities

With regard to the relationship between the MD activities perceived as task demands in a retrospective assessment and those actually realized, the results are worth further consideration. The fact that not the perception of MD activities but the perception of deep-level SD activities and of management activities as task demands was related to realized behavior necessitates explanation. The relationship of the perception of deep-level SD activities as task demands with corroboration might indicate that participants who engage in the effortful corroboration of information across texts also engage in effortful deep-level SD activities, which in the current study were not assessed on a behavioral level. In addition, this finding also could be due to an inability of participants to verbalize MD activities or to successfully apply an activity recognized as important (utilization deficiency; e.g., Bjorklund et al., 1997). MD activities are probably mentally less accessible than deep-level SD activities, facilitating the recognition and verbalization of deep-level SD activities, especially for competent readers (see McNamara, 2011; Mokhtari & Reichard, 2002). The increased perception of MD activities as task demands from the first to the second unit speaks in favor of this interpretation, because the readers seem to learn only over time to express what the MD situation requires them to do. However, the frequency of how often MD activities were perceived as task demands (almost 60% across all units) speaks against it. This finding rather suggests that the perception of MD activities as task demands might have been triggered by our test or by the combination of the test with the assessment of perceived task demands.

The observed mismatch between activities perceived as task demands and realized activities might also be related to a lack of metacomprehension accuracy (e.g., Thiede et al., 2003). In a task model, the to-be-achieved outcome and required activities are represented (Britt et al., 2018; Rouet et al., 2017). Thus, in the task model the standard is represented against which readers monitor their progress (e.g., Winne & Hadwin, 1998). If metacomprehension, that is the monitoring of one's understanding of the texts, is not accurate, no appropriate controlling actions (i.e., multiple-document activities) can be taken, even if they are represented as task demands. So far, there is only few research on metacognition in multiple-document comprehension (e.g., List & Alexander, 2015; Wang & List, 2019), which has not targeted the relation of the task model and metacognition. Thus, further research is necessary to address the interplay of task understanding and metacomprehension in MDC.

The negative relationship of proactive sourcing with the perception of management activities as task demands is consistent with the assumption that proactive sourcing provides a framework that guides the subsequent encoding of text (see Hahnel, Kroehne, et al., 2019). Thus, proactive sourcing might reduce cognitive load (Hahnel, Schoor, et al., 2019) and the need for management activities.

It is worth mentioning that there were within-person level effects of the unit with regard to activities both perceived as task demands and realized. This means that the topic and the specific items of each unit play a role for which activities are conducted during unit processing and for which activities are perceived as task demands after unit processing. In the present study, this was taken into account by including the unit as a control variable in all analyses. For future research, it means that studies should specifically consider how readers deal with several topics and tasks. Moreover, future research might also focus on which features of tasks and topics influence readers' perceptions of task demands and the actual activities they engage in.

4.2 Missing Relationship of Multiple-Document Activities Perceived as Task Demands with Multiple Document Comprehension

In contrast to our expectations, MDC was not related to the perception of MD activities as task demands. If any, the perception of deep-level SD activities as task demands relates to MDC; however, in the current study this effect was not big enough to reach significance. This finding raises the question whether MDC differs from single-text comprehension or whether multiple documents only provide an advanced reading situation. One might wonder whether this could be due to the nature of the MDC test and its items. Yet, the MDC test items were not solvable with only one text, but they required the application of MD activities. Therefore, this question has to be addressed empirically by relating the MDC test to a classic reading comprehension test. Results of an independent study show that although MDC is related to single-text comprehension, it can be considered a different construct (Mahlow et al., 2020).

Assuming that MDC is different from single-text comprehension, this lack of relationship could also be due to an inability of participants to verbalize MD activities or to successfully apply an activity recognized as important (utilization deficiency; e.g., Bjorklund et al., 1997), as discussed in the previous section. Moreover, perhaps MD activities also need to be differentiated into surface-level MD activities, such as the comparison of facts across documents, and deep-level MD activities, such as integrating information across documents to infer new information (see Hagen et al., 2014, for inter-textual elaborations in notes).

4.3 Context Schema Development

The increased perception of MD activities as task demands and the increased probability of showing sourcing behavior in later units suggest that the students may have learned over time what the MD assignments asked of them. In terms of RESOLV, they developed a context schema. This can be seen as a form of test-wiseness (see Millman et al., 1965), but also indicates that an adequate understanding of what multiple documents necessitate can be learned in a relatively short period of time (see Britt & Aglinskas, 2002; Stadtler et al., 2018). This learning might have been fostered by the assessment of the task model after each unit. The task model assignment of the first unit might have served as a self-explanation of the task that also affected the following tasks, as the requirements were comparable across units (see Cerdán et al., 2013). Since university students as a group can be considered good readers as compared to secondary-school students, this self-explanation might have helped them to develop an appropriate context schema. This interpretation is also supported by the finding that there was an increase of MD activities (and a decrease of surface-level SD and management activities) being perceived as task demands from the first to the second assessment. There was no more significant change from the second to the third assessment.

Alternatively, the MDC test might have trained MD reading skills[2], although it is not explicitly conceptualized as training and did not provide feedback. Answering the comprehension items may have created a training effect, since they were not solvable with only one text, stressing the importance of comparison of texts and sources and of the integration of content across documents. However, whether or not this includes the development of a context schema cannot be answered with the present data. Moreover, it is still an open question whether the learning that occurred can be transferred to a different type of MD assignment other than the test used. Yet, the fact that out of all researched indicators of realized behavior and of perceived task demands, the realized behavior of proactive sourcing was the only one hat increased not only from the first to the second unit but also from the second to the third unit, suggests that the participants learned not only through self-explanation during the task demands assessment, but that they also learned during and by working on the test that early attention to sources helps them in their further progress with the test.

It is striking that the change of realized MD activities is more pronounced for sourcing, especially proactive sourcing, and not visible for corroboration. This might have several reasons. First, participants might have understood what the task demanded of them (and reported so), but still failed to act accordingly due to lack of motivation (in a low-stakes test). Proactive sourcing, on the other hand, might have saved them time by making it easier to understand the texts and answering the items (Hahnel, Kroehne, et al., 2019)(Wineburg,

[2] We thank an anonymous reviewer for this thought.

1991). This might not be equally true for corroboration, especially since in the present study corroboration was measured as overall text switches and was not further differentiated (e.g. in spontaneous vs. item-triggered corroboration). A second explanation might be that in case of corroboration participants understood and acted on the task demands, but that they did so in an unobservable manner. They might have learned to pay more attention to commonalities and discrepancies across texts during reading, but did so using their working memory. In this case, we would not be able to see an increase in this activity in logfile data. A third explanation lies in the state of research on (indicators of) sourcing and corroboration: There is much more research on sourcing than on corroboration, including research on how to build indicators based on logfile data. It is also easier to capture sourcing behavior (at least sourcing behavior that uses explicit source information) in logfile data by requiring an event (like pressing a button) in order to access source information. In contrast, it is less clear why participants switched between texts. Therefore, the indicator for corroboration might have captured more noise than the indicators for sourcing. The question of the validity of different ways to build indicators for corroboration has to be addressed in further research.

4.4 Limitations

The present study entails a range of limitations. Firstly, readers' perception of task demands was measured only once at the end of a unit. We chose to focus on this final state of the task model as a first step, since we did not want the reflection on task demands to influence the natural task performance. This is because prior research suggests that a task model assessment before working on the task might change how the task is addressed (Llorens & Cerdán, 2012; McNamara, 2017). While this is an approach in line with the assumptions of RESOLV, the data does not provide information about the development of the task model during the work process nor about the relationship of the task model at the beginning of the task with realized MD behavior and MDC, which are still open questions. Also a matter of future research is the causal relationship between task model and MDC. An adequate task model might lead to a better MDC, but an adequate task model could also be a part of the multiple document competence reflected in MDC test scores. One might also wonder whether the assessment of task demands at the end of the assignment is not just a retrospective self-report of conducted strategies. However, we carefully framed the task demands question to avoid this bias. We did not ask the participants to report the strategies they actually applied, but to describe what actions were needed to be taken. As such, the questions should have prompted the participants to report what they were asked to do and not what they did. We think the data supports this interpretation as the participants factually reported what the task asked them to do, not what activities they actually conducted (Table 4).

Secondly, we analyzed the task model in a test situation with low stakes for the participants. Therefore, they might not have shown their full potential (e.g., Eklöf, 2010; Wise & DeMars, 2005). Examining the effects of task importance (or stakes) on readers' task interpretation and actual behavior is especially important given the demands of multiple document comprehension on effort and strategic behavior. Thirdly, the MDC test was constructed such that participants have low prior content knowledge. Accordingly, effects of prior beliefs and prior knowledge were reduced as much as possible, but in real-life situations they might have a strong impact (see Bråten et al., 2014; Richter & Maier, 2017). For example, with prior knowledge the participants might be more aware of their activities or required MD activities. Moreover, the present results should be replicated with other topics and texts since characteristics of the present ones might have influenced the results. For example, the unit “2134” stood out a bit since it was recognizably fictitious and there were more differences in characteristics of sources compared to other units. This might have impacted especially the sourcing behavior of readers (see Braasch et al., 2012). Finally, the participants in the study were university students of the social sciences and humanities. Therefore, the results are not generalizable beyond this population.

4.5 Conclusion

Taken together, the present study provides new insight into the activities that university students perceive as task demands in a retrospective assessment after they completed a multiple document assignment. Based on Britt et al.'s (2018) RESOLV theory of reading as problem solving, we showed that this perception of task demands is related to differences in realized behavior. More precisely, the perception of deep-level single-document activities as task demands was related to corroboration, and proactive sourcing behavior was negatively related to the perception of management activities as task demands. After subsequent similar assignments, readers more often perceived multiple-document activities as task demands, suggesting learning about the demands of multiple-document assignments. Therefore, readers' perceptions of task demands might be an interesting starting point for designing support for readers who struggle with multiple document comprehension.

References

Afflerbach, P., & Cho, B.-Y. (2009). Identifying and describing constructively responsive comprehension strategies in new and traditional forms of reading. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 69-90). Routledge.

Anmarkrud, Ø., Bråten, I., & Strømsø, H. I. (2014). Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences, 30*, 64-76. https://doi.org/10.1016/j.lindif.2013.01.007

Bjorklund, D. F., Miller, P. H., Coyle, T. R., & Slawinski, J. L. (1997). Instructing children to use memory strategies: Evidence of utilization deficiencies in memory training studies. *Developmental Review, 17*(4), 411-441. https://doi.org/10.1006/drev.1997.0440

Braasch, J. L. G., Rouet, J.-F., Vibert, N., & Britt, M. A. (2012). Readers' use of source information in text comprehension. *Memory & Cognition, 40*(3), 450-465. https://doi.org/10.3758/s13421-011-0160-6

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education, 53*(4), 1207-1217. https://doi.org/10.1016/j.compedu.2009.06.004

Bråten, I., Anmarkrud, Ø., Brandmo, C., & Strømsø, H. I. (2014). Developing and testing a model of direct and indirect relationships between individual differences, processing, and multiple-text comprehension. *Learning and Instruction, 30*, 9-24. https://doi.org/10.1016/j.learninstruc.2013.11.002

Bråten, I., Stadtler, M., & Salmerón, L. (2018). The role of sourcing in discourse comprehension. In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *The Routledge handbook of discourse processes* (pp. 141-166). Routledge/Taylor & Francis Group.

Bråten, I., & Strømsø, H. I. (2010). Effects of task instruction and personal epistemology on the understanding of multiple texts about climate change. *Discourse Processes, 47*(1), 1-31. https://doi.org/10.1080/01638530902959646

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction, 20*(4), 485-522. https://doi.org/10.1207/s1532690xci2004_2

Britt, M. A., & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. In J. R. Kirby & M. J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes* (pp. 276-314). Cambridge University Press.

Britt, M. A., Rouet, J.-F., & Durik, A. M. (2018). *Literacy beyond text comprehension. A theory of purposeful reading*. Routledge.

Cerdán, R., Gilabert, R., & Vidal-Abarca, E. (2013). Self-generated explanations on the question demands are not always helpful. *The Spanish Journal of Psychology, 16.* https://doi.org/10.1017/sjp.2013.45

Cerdán, R., Pérez, A., Vidal-Abarca, E., & Rouet, J.-F. (2019). To answer questions from text, one has to understand what the question is asking: differential effects of question aids as a function of comprehension skill. *Reading and Writing, 32*(8), 2111-2124. https://doi.org/10.1007/s11145-019-09943-w

Cerdán, R., & Vidal-Abarca, E. (2008). The effects of tasks on integrating information from multiple documents. *Journal of Educational Psychology, 100*(1), 209-222. https://doi.org/10.1037/0022-0663.100.1.209

Eklöf, H. (2010). Skill and will: Test-taking motivation and assessment quality. *Assessment in Education: Principles, Policy & Practice, 17*(4), 345-356. https://doi.org/10.1080/0969594X.2010.516569

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906-911.

Francis, D. J., Kulesz, P. A., & Benoit, J. S. (2018). Extending the simple view of reading to account for variation within readers and across texts: The complete view of reading (CVRi). *Remedial and Special Education, 39*(5), 274-288. https://doi.org/10.1177/0741932518772904

Gil, L., Bråten, I., Vidal-Abarca, E., & Strømsø, H. I. (2010). Summary versus argument tasks when working with multiple documents: Which is better for whom? *Contemporary Educational Psychology, 35*(3), 157-173. https://doi.org/10.1016/j.cedpsych.2009.11.002

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., Lee, C. D., Shanahan, C., & Project, R. (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

Hagen, Å. M., Braasch, J. L. G., & Bråten, I. (2014). Relationships between spontaneous note-taking, self-reported strategies and comprehension when reading multiple texts in different task conditions. *Journal of Research in Reading, 37*(S1), S141-S157. https://doi.org/10.1111/j.1467-9817.2012.01536.x

Hahnel, C., Kroehne, U., Goldhammer, F., Schoor, C., Mahlow, N., & Artelt, C. (2019). Validating process variables of sourcing in an assessment of multiple document comprehension. *British Journal of Educational Psychology, 89*(3), 524-537. https://doi.org/10.1111/bjep.12278

Hahnel, C., Schoor, C., Kroehne, U., Goldhammer, F., Mahlow, N., & Artelt, C. (2019). The role of cognitive load for university students' comprehension of multiple documents. *Zeitschrift für Pädagogische Psychologie, 33*(2), 105-118. https://doi.org/10.1024/1010-0652/a000238

Hynd, C., Holschuh, J. P., & Hubbard, B. P. (2004). Thinking like a historian: College students' reading of multiple historical documents. *Journal of Literacy Research, 36*(2), 141-176. https://doi.org/10.1207/s15548430jlr3602_2

Kammerer, Y., Kalbfell, E., & Gerjets, P. (2016). Is this information source commercially biased? How contradictions between web pages stimulate the consideration of source information. *Discourse Processes, 53*(5-6), 430-456. https://doi.org/10.1080/0163853x.2016.1169968

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

Kroehne, U., & Goldhammer, F. (2018). How to conceptualize, represent, and analyze log data from technology-based assessments? A generic framework and an application to questionnaire items. *Behaviormetrika, 45*(2), 527-563. https://doi.org/10.1007/s41237-018-0063-y

Le Bigot, L., & Rouet, J.-F. (2007). The impact of presentation format, task assignment, and prior knowledge on students' comprehension of multiple online documents. *Journal of Literacy Research, 39*(4), 445-470. https://doi.org/10.1080/10862960701675317

Lenhard, W., & Lenhard, A. (2014). *Berechnung des Lesbarkeitsindex LIX nach Björnson* [Computation of the readability index LIX according to Björnson]. http://www.psychometrica.de/lix.html

List, A., & Alexander, P. A. (2015). Examining response confidence in multiple text tasks. *Metacognition and Learning, 10*(3), 407-436. https://doi.org/10.1007/s11409-015-9138-2

List, A., Du, H., & Wang, Y. (2019). Understanding students' perceptions of task assignments. *Contemporary Educational Psychology, 59*, 101801. https://doi.org/10.1016/j.cedpsych.2019.101801

Llorens, A. C., & Cerdán, R. (2012). Assessing the comprehension of questions in task-oriented reading. *Revista de Psicodidáctica, 17*(2). https://www.ehu.eus/ojs/index.php/psicodidactica/article/view/4496

Lorch, R. F., Lorch, E. P., & Klusewitz, M. A. (1993). College students' conditional knowledge about reading. *Journal of Educational Psychology, 85*(2), 239-252.

Mahlow, N., Hahnel, C., Kroehne, U., Artelt, C., Goldhammer, F., & Schoor, C. (2020). More than (single) text comprehension? On university students' understanding of multiple documents. *Frontiers in Psychology, 11*, 562450. https://doi.org/10.3389/fpsyg.2020.562450

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: I. Outcome and process. *British Journal of Educational Psychology, 46*(1), 4-11.

Mason, L., Pluchino, P., & Ariasi, N. (2014). Reading information about a scientific phenomenon on webpages varying for reliability: an eye-movement analysis. *Educational Technology Research and Development, 62*(6), 663-685. https://doi.org/10.1007/s11423-014-9356-3

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review, 19*(2), 113-139. http://dx.doi.org/10.1007/s10648-006-9010-7

McNamara, D. S. (2011). Measuring deep, reflective comprehension and learning strategies: challenges and successes. *Metacognition and Learning, 6*(2), 195-203. https://doi.org/10.1007/s11409-011-9082-8

McNamara, D. S. (2017). Self-explanation and reading strategy training (SERT) improves low-knowledge students' science course performance. *Discourse Processes, 54*(7), 479-492. https://doi.org/10.1080/0163853x.2015.1101328

Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of test-wiseness. *Educational and Psychological Measurement, 25*(3), 707-726. https://doi.org/10.1177/001316446502500304

Mokhtari, K., & Reichard, C. A. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*(2), 249-259. https://doi.org/10.1037/0022-0663.94.2.249

Murphy, P. K., & Alexander, P. A. (2002). What counts? The predictive powers of subject-matter knowledge, strategic processing, and interest in domain-specific performance. *The Journal of Experimental Education, 70*(3), 197-214. https://doi.org/10.1080/00220970209599506

Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99-122). Lawrence Erlbaum Associates.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3), 801-813. https://doi.org/10.1177/0013164493053003024

Richter, T., & Maier, J. (2017). Comprehension of multiple documents with conflicting information: A two-step model of validation. *Educational Psychologist, 52*(3), 148-166. https://doi.org/10.1080/00461520.2017.1322968

Rölke, H. (2012). The ItemBuilder: A graphical authoring system for complex item development. In T. Bastiaens & G. Marks (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2012* (Vol. 2012, pp. 344-353). AACE.

Rouet, J.-F. (2006). *The skills of document use: From text comprehension to Web-based learning*. Erlbaum.

Rouet, J.-F., & Britt, M. A. (2011). Relevance processes in multiple documents comprehension. In M. McCrudden, J. P. Magliano, & G. J. Schraw (Eds.), *Text relevance and learning from text* (pp. 19-52). Information Age.

Rouet, J.-F., Britt, M. A., & Durik, A. M. (2017). RESOLV: Readers' representation of reading contexts and tasks. *Educational Psychologist, 52*(3), 200-215. https://doi.org/10.1080/00461520.2017.1329015

Rouet, J.-F., Favart, M., Britt, M. A., & Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction, 15*(1), 85-106. https://doi.org/10.1207/s1532690xci1501_3

Rouet, J.-F., Rupp, K., Lescarret, C., Steciuch, C., & Britt, M. A. (2020). *Whether, what and how to read: Students' understanding of the reading context determines their information-seeking strategies* [Unpublished manuscript]. University of Poitiers, France.

Schoor, C., Hahnel, C., Artelt, C., Reimann, D., Kröhne, U., & Goldhammer, F. (2020). Entwicklung und Skalierung eines Tests zur Erfassung des Verständnisses multipler Dokumente von Studierenden [Developing and scaling a test of multiple document comprehension in university students]. *Diagnostica, 66*(2), 123-135. https://doi.org/10.1026/0012-1924/a000231

Schoor, C., Hahnel, C., Mahlow, N., Klagges, J., Kroehne, U., Goldhammer, F., & Artelt, C. (2020). Multiple document comprehension of university students: Test development and relations to person and process characteristics. In O. Zlatkin-Troitschanskaia & H. A. Pant (Eds.), *Student Learning Outcomes in Higher Education* (pp. 221-240). Springer.

Snow, C., & the RAND Reading Study Group. (2002). *Reading for understanding. Toward an R&D program in reading comprehension*. RAND.

Stadtler, M., & Bromme, R. (2014). The content–source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In D. N. Rapp & J. L. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 379-402). MIT Press.

Stadtler, M., Bromme, R., & Rouet, J.-F. (2018). Learning from multiple documents: How can we foster multiple document literacy skills in a sustainable way? In E. Manalo, Y. Uesaka, & C. A. Chinn (Eds.), *Promoting spontaneous use of learning and reasoning strategies: Theory, research, and practice for effective transfer* (pp. 46-61). Routledge.

Stadtler, M., Scharrer, L., Brummernhenrich, B., & Bromme, R. (2013). Dealing with uncertainty: Readers' memory for and use of conflicting information from science texts as function of presentation format and source expertise. *Cognition and Instruction, 31*(2), 130-150. https://doi.org/10.1080/07370008.2013.769996

Stadtler, M., Scharrer, L., Skodzik, T., & Bromme, R. (2014). Comprehending multiple documents on scientific controversies: Effects of reading goals and signaling rhetorical relationships. *Discourse Processes, 51*(1-2), 93-116. https://doi.org/10.1080/0163853x.2013.855535

Thiede, K. W., Anderson, M. C. M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66-73. https://doi.org/10.1037/0022-0663.95.1.66

Van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition, 29*(8), 1081-1087. https://doi.org/10.3758/bf03206376

Wang, Y., & List, A. (2019). Calibration in multiple text use. *Metacognition and Learning, 14*(2), 131-166. https://doi.org/10.1007/s11409-019-09201-y

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427-450. https://doi.org/10.1007/bf02294627

Weinstein, C. E., & Mayer, R. E. (1986). The teaching of learning strategies. In M. Wittrock (Ed.), *Handbook of research on teaching* (pp. 315-327). Macmillan.

Wiley, J., Steffens, B., Britt, M. A., & Griffin, T. D. (2014). Writing to learn from multiple-source inquiry activities in history. In P. D. Klein, P. Boscolo, L. C. Kirkpatrick, & C. Gelati (Eds.), *Writing as a learning activity* (pp. 120-148). Brill. https://doi.org/10.1163/9789004265011_007

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*(2), 301-311. https://doi.org/10.1037/0022-0663.91.2.301

Wineburg, S. S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*(1), 73-87. https://doi.org/10.1037/0022-0663.83.1.73

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Lawrence Erlbaum.

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10*(1), 1-17. https://doi.org/10.1207/s15326977ea1001_1

Wolfe, M. B. W., & Goldman, S. R. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction, 23*(4), 467-502. https://doi.org/10.1207/s1532690xci2304_2

Table 1. Units of the multiple document comprehension test and their reading goals.

Unit	Texts	Text length (words)	LIX	# items (# used for test score estimation)	Reading goal
Nothing	2 reviews of the fictitious novel "Nothing" describing content and quality of the novel	723, 562	46.52, 50,96	36 (13)	Please read the texts as if afterwards you had to describe the content of the novel "Nothing" and its quality.
Universe	3 popular science texts on the end of the universe from a physical-cosmological perspective (scenarios of the end, description of forces, report of new empirical data)	455, 464, 448	41.03, 41.31, 44.78	17 (15)	Please read the texts as if afterwards you had to describe how the end of the universe is related to the different forces and what the dark energy has to do with this.
Catalano	2 short biographies on the life of the (fictitious) mafia boss Catalano	644, 584	48.67, 46.13	22 (11)	Please read the texts as if afterwards you had to describe the course of life of Catalano.

Forgiving	3 textbook texts, each presenting a (fictitious) theoretical model on forgiving	853, 586, 828	53.22, 55.05, 54.35	35 (0)	Please read the texts as if afterwards you had to give a presentation in a university course based on these texts.
2134	3 texts on a “historical” event in the year 2134: the arrival of extra-terrestrials on earth. One text is an internal report of an observatory, one an internal governmental report, and one is a political speech	491, 434, 381	50.67, 49.71, 53.32	21 (11)	Please read the texts as if afterwards you had to summarize the events related to the arrival of the extra-terrestrials.
Animals	3 textbook texts each presenting one (fictitious) literature studies approach on how to interpret animals in novels	629, 1057, 451	53.56, 54.81, 51.03	26 (17)	Please read the texts as if afterwards you had to describe how animals in novels could be interpreted.

Note. All texts were written by Schoor, Hahnel, Artelt, et al. (2020). LIX: Readability index according to Lenhard and Lenhard (2014).

Table 2. Types of items and sample items.

Type of item	Requirement	Example
1. Corroboration of information across texts	Information from different texts has to be compared. The information is either directly contained in the text or a simple inference has to be drawn. Referring to Wineburg's (1991) strategy of corroboration.	Do the statements in the three texts agree with regard to the following issues? a) The appraisal of the consequences of forgiving. b) The question whether forgiving depends on culture.
2. Integration of information across texts	Information from different texts has to be combined additively or by means of an inference. Referring to the integrated situation model in the Documents Model Framework (e.g., Britt & Rouet, 2012).	Which statement on the influence of the personality (of the victim), such as the propensity to retaliation, decisiveness, or empathy, on forgiving is correct based on the three texts? • Some personality characteristics influence forgiving directly, whereas for others an indirect relationship is assumed. • Personality characteristics influence forgiving indirectly via the motivation to forgive.

		• It is often assumed that personality characteristics influence forgiving; however, the research results so far do not speak in favor of this assumption. • Forgiving is fundamentally influenced by personality characteristics.
3. Comparison of sources and source evaluations across texts	Sources of single texts have to be judged and compared across texts / sources. Referring to the intertext model in the Documents Model Framework (e.g., Britt & Rouet, 2012).	Are the following statements about the works described in the texts correct? a) The differences in the described works are probably due to the scientific progress. b) The works described in the texts seem to be conducted based on the perspective of different scientific domains.
4. Comparison of source-content links across texts	Information has to be represented with its source. These source-content links have to be compared across texts. Referring to the documents model in the Documents Model Framework (e.g., Britt & Rouet, 2012).	Compare the three dimensions of factors influencing forgiving according to Thomsen et al. with the process model by Shavelton and van den Bechele. Which statement is correct?

	• All factors influencing forgiving in the model by Shavelton and van den Bechele can be classified into the dimensions according to Thompsen et al., but not all dimensions according to Thompsen et al. can be assigned to one or more phases of the model by Shavelton and van den Bechele. • All dimensions according to Thompsen et al. can be assigned to one or more phases of the model by Shavelton and van den Bechele but not all factors influencing forgiving in the model by Shavelton and van den Bechele can be classified into the dimensions according to Thompsen et al. • Both all dimensions according to Thompsen et al. can be assigned to one or more phases of the model by Shavelton and van den Bechele, and all factors influencing forgiving in the model by Shavelton and van den Bechele can be

	classified into the dimensions according to Thompsen et al. • Neither can all dimensions according to Thompsen et al. be assigned to one or more phases of the model by Shavelton and van den Bechele nor can all factors influencing forgiving in the model by Shavelton and van den Bechele be classified into the dimensions according to Thompsen et al.

Table ***3***. Coding scheme for the analysis of the activities perceived as task demands.

Coding category	description	Fre-quency	Cohen's κ
surface-level single-document activities	memorize information from the texts, list facts, sort facts, attend to details, read the texts several times, look up information in the texts, highlight important information, make notes / comments, summarize, make a table, find titles for paragraphs	50.4%	.98
deep-level single-document activities	think / reflect during reading, imagery, critical thinking, separate relevant from irrelevant content, keep questions in mind	24.2%	.79
multiple-document activities	compare information across texts, find commonalities and differences across texts, relate texts to each other, combine information from the texts in order to infer new information, integrate information across texts, keep in mind which text says what, differentiate between texts, judge / evaluate the source / the differences between the sources, attribute intentions / competence to the source, detect biases of the sources, decide for one position, (re-) construct an own picture of the topic, find the / one's own truth, find an objective version	59.9%	.85
management activities	be fast, take enough time, keep the time in mind, have breaks, concentrate, pay a high attention, don't get distracted	26.1%	.81

Table 4. Examples of coded answers.

	Answer	SL	DL	MD	MA
1	"[…] For the content summary, one had to read both reviews and to select the information that was consistent in both texts" (PB01021 for the unit "Nothing")	no	no	yes	no
2	"It was about 3 (this time personal) texts, 2 "diaries" and one speech. The situation of the Europeans, Africans, and Americans after seeing a UFO was compared. Again, you had to carefully read and understand the texts in order to solve the following tasks." (PB01105 for the unit "2134")	no	no	no	no
3	"You had to read the three texts one after another and read all important main information out that concern the aliens, their arrival, as well as existing nations and their reactions." (PB01011 for the unit "2134")	no	yes	no	no
4	"In order to solve such a task correctly you had to combine and relate. Because the three were so different text forms, it was more difficult to extract the most important [information] than in the previous texts [= units]. You had to concentrate strongly in order to solve the task correctly." (PB01012 for the unit "2134")	no	yes	yes	yes
5	"In order to find out the correctness of the biographical story of the protagonist, you first had to compare the two texts, infer further information from them and filter more information from both texts. In order to solve such a task correctly, you should keep in mind the information of the first text during the careful reading of	yes	yes	yes	no

the second text, and you should highlight important dates and parts of the text." (PB02012 for the unit "Catalano")

Note. "Reading" was considered a basic requirement of the tasks, and therefore not included in any category. SL = surface-level single-document activities; DL = deep-level single-document activities; MD = multiple-document activities; MA = management activities.

Table 5. Intercorrelations of indicators of realized multiple-document activities.

	(1)	(2)	(3)
(1) Corroboration		.28***	.63***
(2) Proactive sourcing	-.00		.71***
(3) Repeated sourcing	.25***	.27***	

Note. * $p < .05$; ** $p < .01$; *** $p < .001$. Within-level (unit-level) correlations below the diagonal, between-level (person level) correlations above the diagonal.

Table 6. Prediction (in odds ratios with standard error in parentheses) of activities perceived as task demands by indicators of corroboration and sourcing in Model 1.

	Corroboration	Proactive sourcing	Repeated sourcing
Surface-level single-document activities	1.01 (0.01)	0.75 (0.20)	1.03 (0.28)
Deep-level single-document activities	1.02 (0.01)*	0.76 (0.22)	0.87 (0.24)
Multiple-document activities	1.02 (0.01)	1.61 (0.43)	0.79 (0.21)
Management activities	1.00 (0.02)	0.46 (0.16) ***	0.92 (0.31)

Note. * $p < .05$; ** $p < .01$; *** $p < .001$. On the unit level, both activities perceived as task demands and realized activities were additionally predicted by the unit.

Table 7. Correlation of activities perceived as task demands with MDC in Model 2.

	MDC
Surface-level single-document activities	.01 (0.08)
Deep-level single-document activities	.17 (.09)
Multiple-document activities	.08 (0.08)
Management activities	-.13 (0.08)

Note. * $p < .05$; ** $p < .01$; *** $p < .001$. On the unit level, activities perceived as task demands were predicted by the unit. Standardized coefficients with standard errors in parentheses. MDC = multiple document comprehension.

Table 8. Within-level effects (odds ratios with standard errors) of the position of the unit (dummy variables with reference position second unit) predicting activities perceived as task demands in Model 3.

	Position = 1	Position = 3
Surface-level single-document activities	2.28 (0.44) **	0.84 (0.18)
Deep-level single-document activities	1.08 (0.23)	0.97 (0.20)
Multiple-document activities	0.63 (0.12) **	0.94 (0.18)
Management activities	2.96 (0.71) **	1.00 (0.24)

Note. * $p < .05$; ** $p < .01$; *** $p < .001$. Activities perceived as task demands were additionally predicted by the unit. Two dummy variables coded difference from position as second unit.

Table 9. Within-level effects (standardized coefficients with standard errors in parentheses) of the position of the unit (dummy variables with reference position second unit) predicting indicators of corroboration and sourcing in Model 4.

	Position = 1	Position = 3
Corroboration	-.01 (0.03)	-.06 (0.03)
Proactive sourcing	-.23 (0.03) ***	.07 (0.03) *
Repeated sourcing	-.11 (0.03) **	.01 (0.04)

Note. * $p < .05$; ** $p < .01$; *** $p < .001$. Realized activities were additionally predicted by the unit. Two dummy variables coded difference from position as second unit.

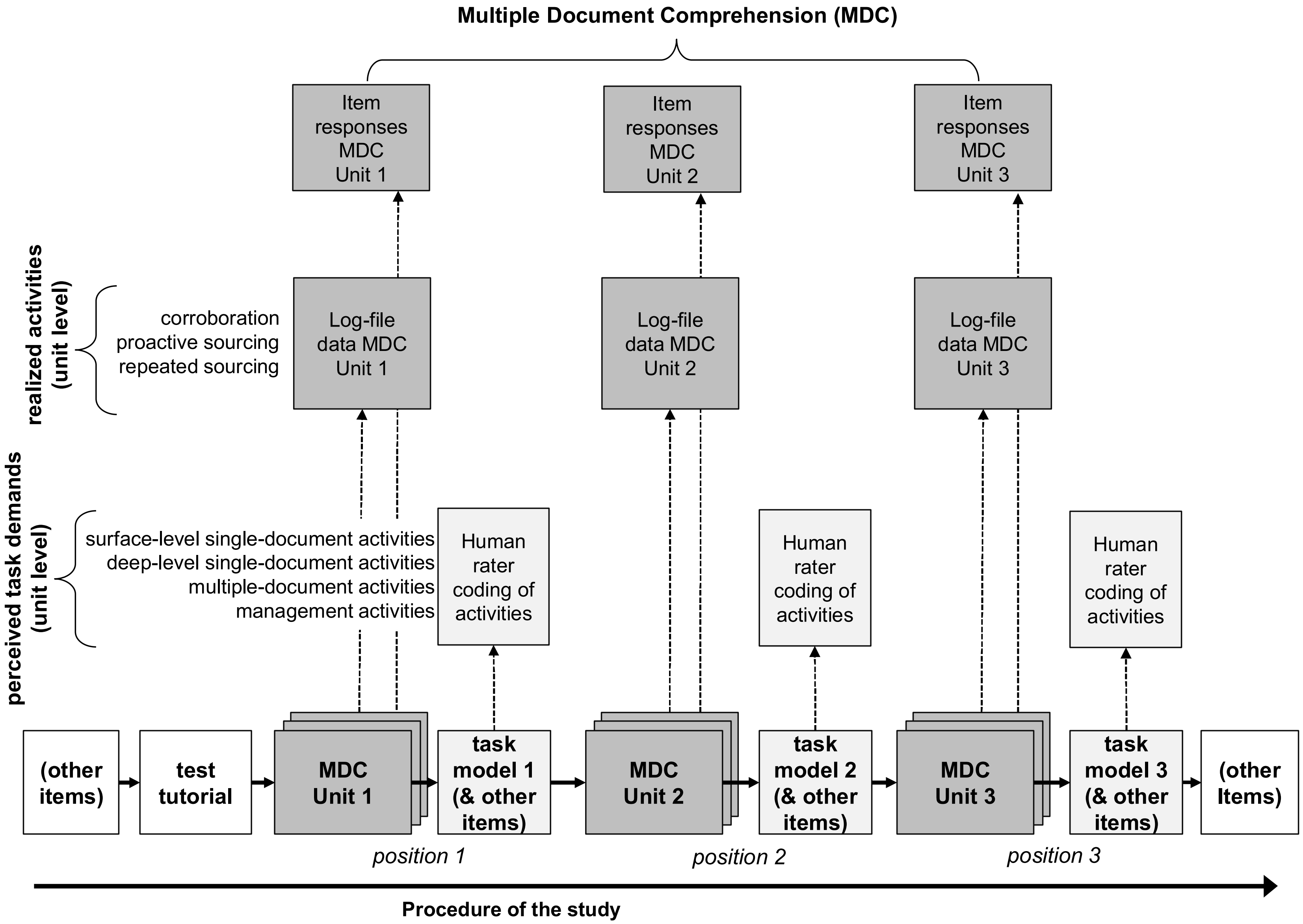
Figure Captions

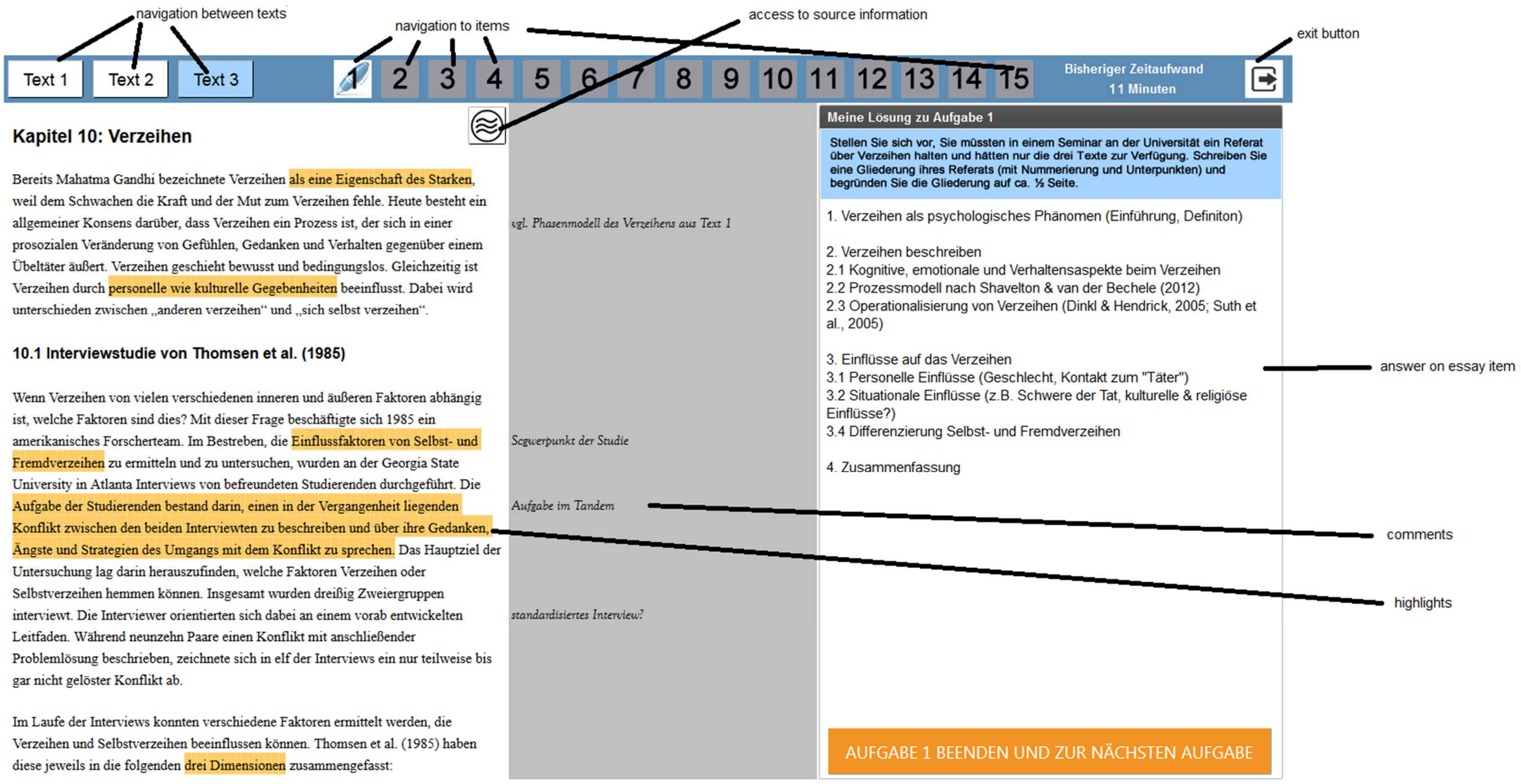
Figure 1. Overview of the procedure of the study (bold arrows) and data sources (dashed arrows). MDC = multiple document comprehension.

Figure 2. Screenshot of the multiple document comprehension test environment.

Figure 3. Models specified for testing Hypotheses 1-4. The circle around level-2 variables symbolizes a random intercept.

Figure 4. Mean probability to mention different categories of activities as task demands by position of the unit (without controlling for unit, error bars represent 1.96 standard errors [i.e., the 95% confidence interval]).


Multiple Document Comprehension (MDC)
Item responses MDC Unit 1
Item responses MDC Unit 2
Item responses MDC Unit 3
realized activities (unit level)
corroboration
proactive sourcing
repeated sourcing
Log-file data MDC Unit 1
Log-file data MDC Unit 2
Log-file data MDC Unit 3
perceived task demands (unit level)
surface-level single-document activities
deep-level single-document activities
multiple-document activities
management activities
Human rater coding of activities
Human rater coding of activities
Human rater coding of activities
(other items)
test tutorial
MDC Unit 1
task model 1 (& other items)
MDC Unit 2
task model 2 (& other items)
MDC Unit 3
task model 3 (& other items)
(other Items)
position 1
position 2
position 3
Procedure of the study



Kapitel 10: Verzeihen

Bereits Mahatma Gandhi bezeichnete Verzeihen als eine Eigenschaft des Starken, weil dem Schwachen die Kraft und der Mut zum Verzeihen fehle. Heute besteht ein allgemeiner Konsens darüber, dass Verzeihen ein Prozess ist, der sich in einer prosozialen Veränderung von Gefühlen, Gedanken und Verhalten gegenüber einem Übeltäter äußert. Verzeihen geschieht bewusst und bedingungslos. Gleichzeitig ist Verzeihen durch personelle wie kulturelle Gegebenheiten beeinflusst. Dabei wird unterschieden zwischen „anderen verzeihen" und „sich selbst verzeihen".

10.1 Interviewstudie von Thomsen et al. (1985)

Wenn Verzeihen von vielen verschiedenen inneren und äußeren Faktoren abhängig ist, welche Faktoren sind dies? Mit dieser Frage beschäftigte sich 1985 ein amerikanisches Forscherteam. Im Bestreben, die Einflussfaktoren von Selbst- und Fremdverzeihen zu ermitteln und zu untersuchen, wurden an der Georgia State University in Atlanta Interviews von befreundeten Studierenden durchgeführt. Die Aufgabe der Studierenden bestand darin, einen in der Vergangenheit liegenden Konflikt zwischen den beiden Interviewten zu beschreiben und über ihre Gedanken, Ängste und Strategien des Umgangs mit dem Konflikt zu sprechen. Das Hauptziel der Untersuchung lag darin herauszufinden, welche Faktoren Verzeihen oder Selbstverzeihen hemmen können. Insgesamt wurden dreißig Zweiergruppen interviewt. Die Interviewer orientierten sich dabei an einem vorab entwickelten Leitfaden. Während neunzehn Paare einen Konflikt mit anschließender Problemlösung beschrieben, zeichnete sich in elf der Interviews ein nur teilweise bis gar nicht gelöster Konflikt ab.

Im Laufe der Interviews konnten verschiedene Faktoren ermittelt werden, die Verzeihen und Selbstverzeihen beeinflussen können. Thomsen et al. (1985) haben diese jeweils in die folgenden drei Dimensionen zusammengefasst:

vgl. Phasenmodell des Verzeihens aus Text 1

Scgwerpunkt der Studie

Aufgabe im Tandem

standardisiertes Interview?

Meine Lösung zu Aufgabe 1

Stellen Sie sich vor, Sie müssten in einem Seminar an der Universität ein Referat über Verzeihen halten und hätten nur die drei Texte zur Verfügung. Schreiben Sie eine Gliederung ihres Referats (mit Nummerierung und Unterpunkten) und begründen Sie die Gliederung auf ca. ½ Seite.

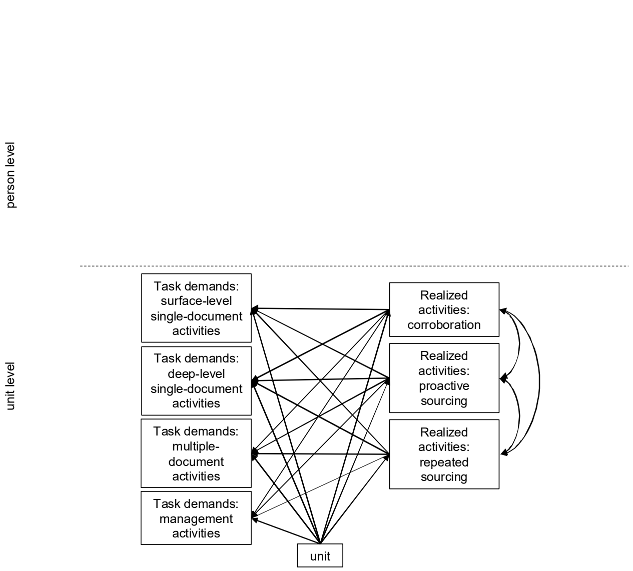
1. Verzeihen als psychologisches Phänomen (Einführung, Definiton)

2. Verzeihen beschreiben
2.1 Kognitive, emotionale und Verhaltensaspekte beim Verzeihen
2.2 Prozessmodell nach Shavelton & van der Bechele (2012)
2.3 Operationalisierung von Verzeihen (Dinkl & Hendrick, 2005; Suth et al., 2005)

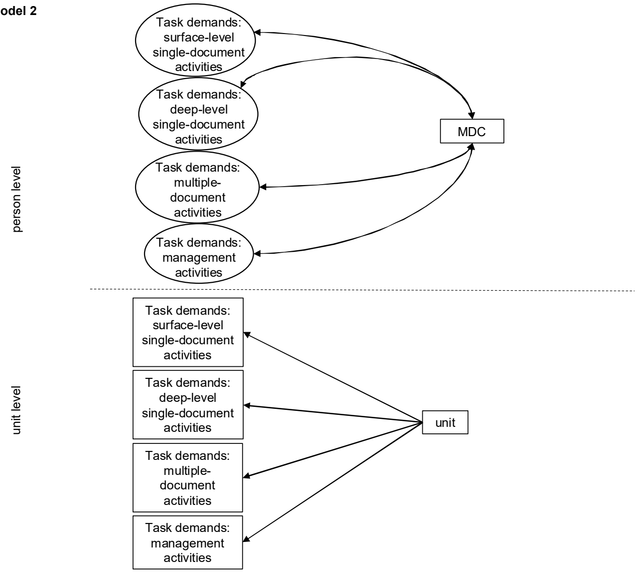
3. Einflüsse auf das Verzeihen
3.1 Personelle Einflüsse (Geschlecht, Kontakt zum "Täter")
3.2 Situationale Einflüsse (z.B. Schwere der Tat, kulturelle & religiöse Einflüsse?)
3.4 Differenzierung Selbst- und Fremdverzeihen

4. Zusammenfassung

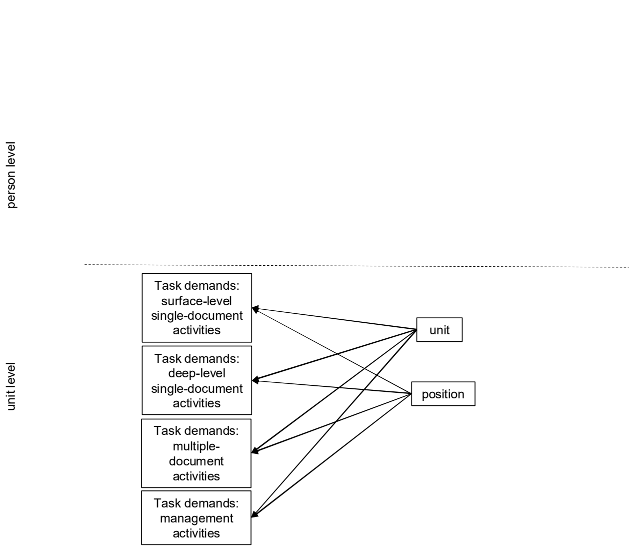
Model 1


person level
unit level
Task demands: surface-level single-document activities
Task demands: deep-level single-document activities
Task demands: multiple-document activities
Task demands: management activities
Realized activities: corroboration
Realized activities: proactive sourcing
Realized activities: repeated sourcing
unit

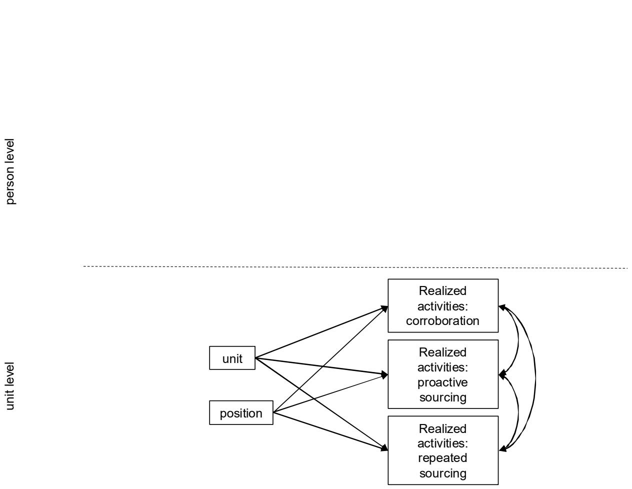
Model 2

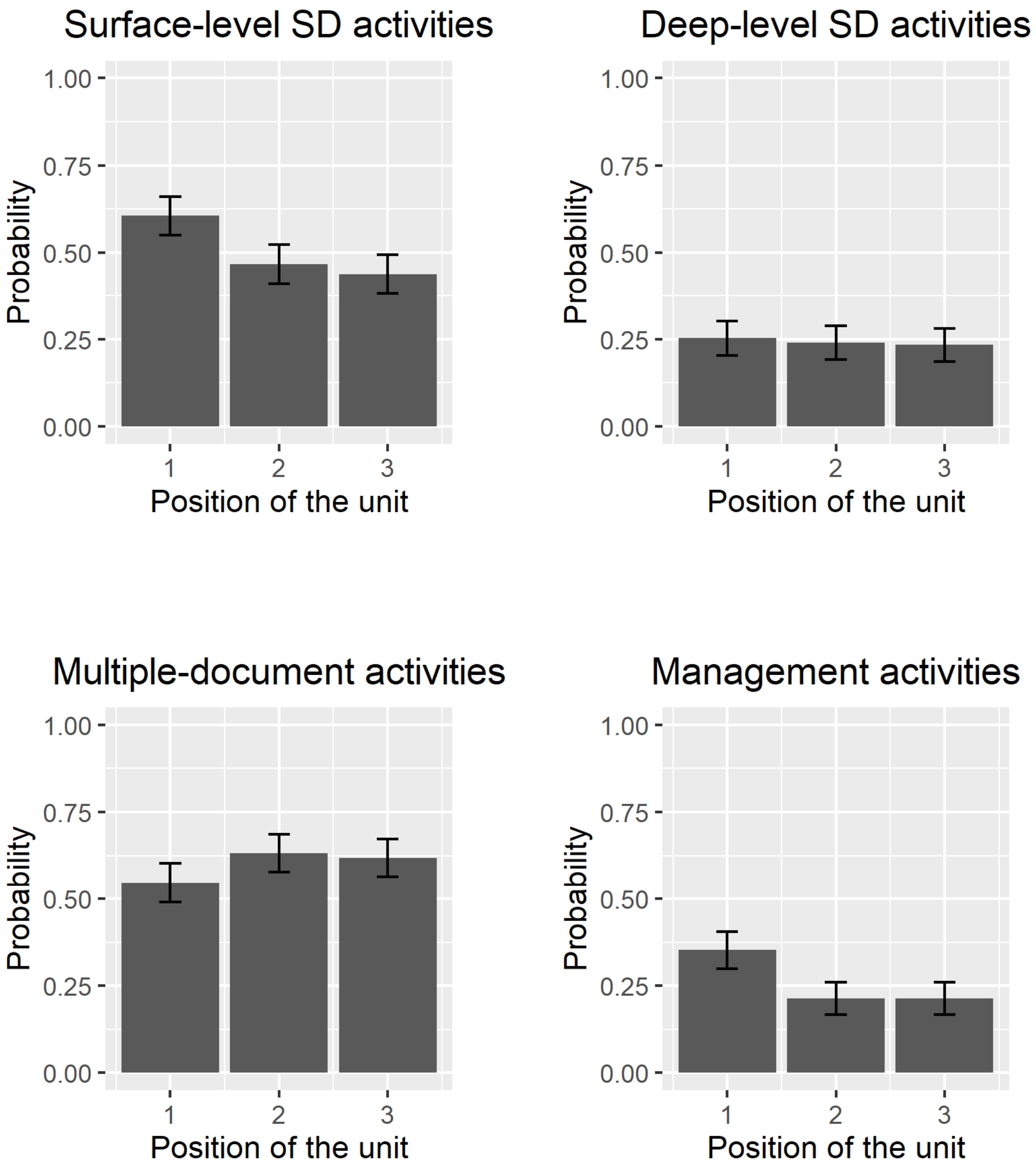

person level
unit level
Task demands: surface-level single-document activities
Task demands: deep-level single-document activities
Task demands: multiple-document activities
Task demands: management activities
MDC
Task demands: surface-level single-document activities
Task demands: deep-level single-document activities
Task demands: multiple-document activities
Task demands: management activities
unit

Model 3


person level
unit level
Task demands: surface-level single-document activities
Task demands: deep-level single-document activities
Task demands: multiple-document activities
Task demands: management activities
unit
position

Model 4


person level
unit level
unit
position
Realized activities: corroboration
Realized activities: proactive sourcing
Realized activities: repeated sourcing


Surface-level SD activities
Deep-level SD activities
Multiple-document activities
Management activities
Probability
1.00
0.75
0.50
0.25
0.00
1
2
3
Position of the unit