The role of cognitive load in university students' comprehension of multiple documents

Carolin Hahnel
DIPF | Leibniz Institute for Research and Information in Education, Center for International Student Assessment (ZIB)
Address: Rostocker Straße 6, 60323 Frankfurt, Germany
Email: hahnel@dipf.de

Cornelia Schoor
University of Bamberg
Address: Markusplatz 3, 96047 Bamberg, Germany
Email: cornelia.schoor@uni-bamberg.de

Ulf Kröhne
DIPF | Leibniz Institute for Research and Information in Education
Address: Rostocker Straße 6, 60323 Frankfurt, Germany
Email: kroehne@dipf.de

Frank Goldhammer
DIPF | Leibniz Institute for Research and Information in Education, Center for International Student Assessment (ZIB)
Address: Rostocker Straße 6, 60323 Frankfurt, Germany
Email: goldhammer@dipf.de

Nina Mahlow
Leibniz Institute for Educational Trajectories (LIfBi)
Address: Wilhelmsplatz 3, 96047 Bamberg, Germany
Email: nina.mahlow@lifbi.de

Cordula Artelt
Leibniz Institute for Educational Trajectories (LIfBi), University of Bamberg
Address: Wilhelmsplatz 3, 96047 Bamberg, Germany
Email: cordula.artelt@lifbi.de

Acknowledgements

The reported study was funded by the German Federal Ministry of Education and Research, funding number 01PK15008, within the research program of KoKoHs ("Kompetenzmodelle und Instrumente der Kompetenzerfassung im Hochschulsektor – Validierungen und methodische Innovationen"). The data were assessed and analyzed as part of the MultiTex project ("Prozessbasierte Diagnostik des Textverstehens mit multiplen Dokumenten"). The responsibility for the content of this publication lies with the authors. We want to thank two anonymous reviewers for their valuable insights and constructive comments as well as Holly A. Nyx for proofreading the paper.

Post print version of the article: 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://dx.doi.org/10.1024/1010-0652/a000238

Die Rolle kognitiver Belastung für das Verständnis multipler Dokumente von Studierenden

Die Studie untersucht das Belastungserleben (*Cognitive Load*) von Studierenden beim Bearbeiten von Aufgaben, die das Verstehen multipler Dokumente erfordern (*Multiple Document Comprehension*, MDC). Es wurde geprüft, inwiefern die wahrgenommene Aufgabenschwierigkeit und die mentale Anstrengung von 310 Studierenden durch aufgabenspezifische Eigenschaften, individuelle Merkmale und ihr Bearbeitungsverhalten bestimmt werden und darüber hinaus MDC erklären. Für die Aufgabenschwierigkeit wurde gezeigt, dass sie mit der Dokumentenanzahl, der Textlänge, dem Studienniveau sowie der Berücksichtigung von Quellen in Zusammenhang steht. Die mentale Anstrengung wurde durch die Textlänge, das Studienniveau und Bearbeitungszeiten vorhergesagt. Unter Einschluss dieser Variablen als Kovariaten war das Belastungserleben inkrementell prädiktiv für MDC. Die Ergebnisse werden dahingehend diskutiert, wie Arbeitsgedächtnisressourcen den Prozess des Verstehens multipler Dokumente gestalten können.

Schlüsselwörter

Multiple Dokumente; Cognitive Load; Mental Load und Mental Effort; Bearbeitungsverhalten; Logdaten

The role of cognitive load in university students’ comprehension of multiple documents

Abstract

The study investigates the cognitive load of students working on tasks that require the comprehension of multiple documents (Multiple Document Comprehension, MDC). In a sample of 310 students, perceived task difficulty (PD) and mental effort (ME) were examined in terms of task characteristics, individual characteristics, and students’ processing behavior. Moreover, it was investigated if PD and ME can still contribute to MDC while controlling for these variables. The perceived difficulty of the task was shown to be related to the number of documents, text length, study level, and sourcing. Mental effort was predicted by text length, study level, and processing time. When including these variables as covariates, cognitive load was incrementally predictive of MDC. The results are discussed in terms of how working memory resources can shape the process of comprehending multiple documents.

Keywords

Multiple documents; cognitive load; mental load and mental effort; processing behavior; log files.

1. Introduction

During learning, the use of information from different sources is an enormous advantage for creating a broad and comprehensive knowledge base (Britt & Rouet, 2012). Because different documents might provide unique, similar or even contradictory perspectives, the skills to compare and integrate content across documents and to evaluate their information with respect to their source are indispensable for achieving a comprehensive understanding of a topic or area of knowledge (Anmarkrud, Bråten, & Strømsø, 2014). As is evident in German educational standards, it is often expected that students who achieved the qualifications to gain entrance to university are skilled in multiple document comprehension (MDC). Yet, empirical studies point out that students often show difficulties in dealing with multiple documents appropriately (e.g., Britt & Aglinskas, 2002; Wiley et al., 2009).

Contributing to the question of why students experience problems when working with multiple documents, the present study focuses on the role of their experience of cognitive load. The requirements of MDC (Britt & Rouet, 2012) are most likely to place increased cognitive demands on readers, as they need to hold and process multiple interacting elements in working memory (WM) at the same time (Sweller, 2010). If cognitive demands then exceed the limited resources of WM (Sweller, Ayres, & Kalyuga, 2011), MDC can be severely impaired. We investigated the extent to which individual differences in MDC are associated with cognitive load by examining factors that might determine the individual amount of cognitive load and affect the students' ability in MDC. Two aspects of cognitive load, the perceived difficulty of tasks (PD) together with the mental effort (ME) students invested in task processing were examined (Paas, 1992); as determinants, we investigated the characteristics of tasks (e.g., number of documents), individuals (e.g., study level), and behaviors shown during task processing (e.g., attention to sources indicating sourcing; Wineburg, 1991).

1.1 Cognitive demands of MDC

To achieve deep comprehension when studying multiple documents, students are required to create a coherent and comprehensive mental structure that represents the views within documents, relationships between them, and the connections of views to their sources (documents model framework; Britt & Rouet, 2012). The 'documents model framework' integrates two sub-levels of representation—the intertext model and an integrated mental model of the situation or phenomena described across texts. The intertext model includes information about the document sources (e.g., author, worldviews, intentions); the integrated mental model reflects a representation of document contents that requires students to comprehend and evaluate arguments within the documents and create cross-document links.

The documents model framework proposes an ideal mental representation of information, which is achieved once students have developed a comprehensive understanding of the breadth, depth and connectedness of information within documents. It can be assumed that attempting to reach this comprehensive understanding places extensive cognitive demands on students, as information needs to be compared and integrated across documents and represented as interpreted by a particular source. If readers cannot handle the amount of information processed simultaneously in WM, comprehension will be impaired according to Cognitive Load Theory (CLT; Sweller et al., 2011). The perceived cognitive load is primarily determined by the level of element interactivity which refers to the number of unique, but interrelated elements held simultaneously in WM. Low element interactivity means that learners can process elements sequentially since there are no dependencies between them (e.g., learning vocabulary); high element interactivity implies the need to process interlinking elements simultaneously (e.g., solving equations; Sweller, 2010). Although reading processes cannot be defined clearly in terms of element interactivity, it is assumed that element interactivity is a prominent feature of written learning materials that require comprehension (Chen et al., 2017).

1.2 Factors affecting cognitive load and MDC

CLT is based on central assumptions about human cognitive architecture, which consists of a virtually unlimited long-term memory and a capacity-limited WM (Paas & Sweller, 2012). The WM capacity limits apply in particular for novel information gathered from the environment; they disappear when the WM deals with familiar information organized in cognitive schemas. Schemas incorporate multiple elements of information grouped into single elements and work as automated rules for task processing (Paas, 1992). By supporting decision-making, automated schemas require fewer cognitive resources, as they reduce the number of elements to be held in WM (Sweller, 2010). Accordingly, cognitive load is reduced when new information can be linked with existing schemas.

Cognitive load is often distinguished as being intrinsic (inherent to the learning material) or extraneous (irrelevant to learning and due to material presentation). There is an ongoing debate on germane load as a third type, resulting from beneficial learning activities (Klepsch, Schmitz, & Seufert, 2017; Korbach, Brünken, & Park, 2018). Empirically though, germane load cannot often be clearly differentiated, which is why it was suggested to only assume germane resources are used for schema acquisition and linking comprehension of information with already existing knowledge structures (Choi, van Merriënboer, & Paas, 2014). A more general differentiation is made between mental load and mental effort (Paas, Tuovinen, Tabbers, & Van Gerven, 2003). Mental load is the load that the characteristics of a task impose on learners, including intrinsic and extraneous load. It indicates a learner's estimation of the demands that a task might place on their cognitive capacity and is often investigated in terms of perceived difficulty of tasks (PD). Mental effort (ME) is the amount of cognitive effort that learners actually allocate to the task processing during action on the task. Since it can be seen as a learner's reaction to meet the expected demands, ME is a prerequisite for instructional settings in which comprehension is affected. If ME needs to be sustained over a longer period of time, performance may be compromised due to already exhausted resources (WM depletion; Chen, Castro-Alonso, Paas, & Sweller, 2018). In the following, we briefly introduce factors that might affect the PD and ME of students studying multiple documents.

Task characteristics. The implications of CLT provide guidance for the design of educational material (Leppink & van den Heuvel, 2015). For example, a redundant (redundancy effect) or split representation of information (split attention effect) should be avoided. In this respect, learning with multiple documents seems inadequate. The distribution of information over several documents introduces interruptions to the reading process and requires learners to keep their mental representation active while deciding about whether and in which order to access other sources and integrate information into their existing overall representation (DeStefano & LeFevre, 2007). As source information has to be represented as well, related elements also need to be held in WM, increasing the number of elements proportional to the number of sources and overstressing cognitive load. Yet, other task characteristics might promote deep processing and have a cognitive easing effect. For example, students apply deep processing strategies more often if they do a pre-task related to the topic (Rouet, Ros, Goumi, Macedo-Rouet, & Dinet, 2011). Writing essays can also engage students in deep processing (Britt & Sommer, 2004) by reducing element interactivity through the act of organizing knowledge or serving as an external resource that facilitates solving subsequent tasks (distributed cognition; Hutchins, 1995).

Individual characteristics. One of the most important factors determining the level of cognitive load is expertise or prior knowledge (e.g., Le Bigot & Rouet, 2007). Expertise means that learners have acquired cognitive schemas throughout their level of knowledge over a period of learning (Paas & Sweller, 2012) that support the reduction of element interactivity. Accordingly, students with favorable characteristics, such as prior knowledge, performed better on problem-solving tasks and reported less cognitive load than students with less favorable characteristics (Scheiter, Gerjets, Vollmann, & Catrambone, 2009). However, the use of instructional designs that are suitable for novices can also result in an increase of the cognitive load for experts (expertise reversal effect; e.g., Chen et al., 2017). In this regard, indicators of general performance, such as graduation grades and academic experience, can be of interest. In Germany, graduation grades ("Abiturnote") are an aggregation of subject-specific grades over a period of time assessed by several teachers. They are often considered to represent general cognitive abilities and skills, and have been shown to be highly predictive of academic success (Trapmann, Hell, Weigand, & Schuler, 2007). Concerning academic experience, Kobayashi (2009) found that third year university students outperformed first year students in the comprehension of intertextual relations between documents, and concluded that academic experience fosters students' skills to process and integrate information across texts.

Behavioral factors. Behavioral factors are reflections of how learners interact with particular tasks. In reading research, indicators of time spent on a task, reading strategies and annotations are often considered. When concerned with the time spent on tasks, we already know that skilled readers read faster in simple reading tasks than less skilled readers, as their reading relies on high automatization (Perfetti, 2007). Fast task processing combined with less effort is considered to indicate low cognitive load (Choi et al., 2014). However, tasks that require the controlled processing of information are associated with longer processing times (Goldhammer et al., 2014). With respect to reading strategies, Wineburg (1991) identified MDC-specific strategies by comparing historians and high school students making sense of historical documents. He found differences in the acts of comparing documents to identify consistencies and discrepancies (corroboration) and attending to source information to activate text schemas (sourcing). There is a growing body of research that suggests students have difficulties in applying these strategies appropriately (Britt & Rouet, 2012). For example, students often fail to seek out to sources spontaneously (Britt & Aglinskas, 2002) and rarely consider them for evaluating the reliability of information (Wiley et al., 2009). Concerning annotations, note-taking can work as a memory-offloading strategy and facilitate comprehension by transferring information from memory to an external location (Moos, 2009). Accordingly, Kobayashi (2009) and Hagen, Braasch and Bråten (2014) demonstrated positive effects of note-taking on the comprehension of intertextual relationships.

1.3 Hypotheses

Studying multiple documents can be a cognitively challenging task that university students need to accomplish. The students' perceived cognitive load is likely to play a critical role in the creation of a comprehensive documents model. The present study therefore examines the characteristics of tasks, individuals, and their behaviors as determinants of PD, ME and MDC.

For task characteristics, we investigated the number of provided documents, the total combined text length of documents, and the requirement to complete a pre-task before completing MDC tasks. We expected that the PD would increase by the number of documents (H1) and the total length of documents (H2), but decrease when completing a pre-task (H3). ME was not considered for explanation since it can be assumed that students rate their efforts based on their individual engagement. With respect to the individual characteristics and behavioral factors, three factors were examined in how they relate to the aspects of cognitive load: students' study level, graduation grades and their behavior of working with multiple documents. For behavioral variables, we investigated processing times, the MD-specific strategies corroboration and sourcing, and note-taking. Students were expected to be more successful in solving MDC tasks correctly if they applied MDC-specific strategies (H4a) and note-taking (H4b).

Lastly, although cognitive load and MDC are interdependent, we investigated PD and ME as predictors of MDC to investigate their relationship. We expected the probability of solving MDC tasks correctly to be negatively associated with students' PD and positively with their ME (H5). We further examined possible effects of WM depletion, which was expected to be reflected in a decrease in performance over time (H6; Chen et al., 2018).

2. Methodology

2.1 Participants

An ad-hoc sample of 310 students (79.4% female) from two German universities was investigated (the same sample as in Schoor et al., 2018). They were enrolled in different social sciences and humanities programs (68.7% Bachelor, 31.3% Master) and aged 18 to 34 years (M = 21.44, SD = 2.72). The participants received an expense allowance of 20,-€ and also took part in a lottery to win a voucher for a tablet computer.

2.2 Measures

An overview of all predictor variables is provided in Tab.1. Unintentional missing values occurred rarely (< 1%).

«Include Tab.1, Fig.1»

Multiple Document Comprehension. The computer-based MDC test of Schoor et al. (2018) was used to capture MDC as the cross-disciplinary ability of students to construct a comprehensive representation of a particular topic from various sources of information. The test assesses the skill to compare, integrate and link the content and sources of largely credible documents (i.e., without searching and selecting multiple documents). It is structured in 'units', consisting of documents and items that refer to these documents (i.e., 2-3 texts with up to 16 items; Fig.1). As described in Schoor et al. (2018), although 174 items within six units were developed, only a final set of 67 items from five units was selected due to their psychometric properties (e.g., the sixth unit was excluded due to the effects of differential item functioning penalizing the ability estimation of men compared to women). The remaining items had a single-choice format (31 items with two response alternatives; 36 items with four response alternatives) and were shown to fit a Rasch model sufficiently (EAP reliability=.69). For a correct item response, information from at least two documents in a unit had to be considered. Students' dichotomously scored responses served as a dependent variable for the present study (correct response rates: 16.3% to 90.7%).

The unit contents address topics from different domains (e.g., science, literary studies), and vary in the number of documents, total document lengths and the requirement of essay-writing (Tab.1). To measure MDC as independently as possible of students' prior knowledge, the contents have been mostly invented (except for the unit "universe"). Although the units display different texts and items, they are comparable in structure and functionalities (Fig.1). A unit starts by informing students about the number of documents and items and setting a reading goal (e.g., read the texts as if you had to give a presentation for a seminar). During the study, students could navigate freely between the texts and items, except in units including an essay. In this case, the students were required to write the essay before they got access to the items. Each text page provided a button that opened an additional popup dialog presenting information about the document's source. Students could highlight text passages, set comments next to the text, and received visual feedback on their processing time and task progress. No time restrictions were administered.

«Include Tab.2»

Processing behavior. The behavioral indicators were constructed from log-file data recorded during unit processing. Unit processing times describe the time interval from the start to the end of a unit. The number of transitions between texts indicated corroboration (Wineburg, 1991). A dichotomous variable of whether or not the students accessed all source information available indicated sourcing (Bråten, Stadtler, & Salmerón, 2018).

Students' notes were coded according to whether or not each note included (1) comprehension monitoring, (2) intratextual paraphrasing, (3) intratextual overview, (4) intertextual paraphrasing, (5) intertextual comparison, and (6) source information (Tab.3). These categories were based on strategies of learning from (single) texts and multiple documents (e.g., Weinstein & Mayer, 1983; Wineburg, 1991). All in all, 4,871 comments were coded by two independent trained coders. The overall interrater reliability was Cohen's κ=.68, which we have interpreted as substantial agreement (Landis & Koch, 1977). For the present analyses, the number of comments assigned to each category was dichotomized per participant and unit.

«Include Tab.3»

Perceived task difficulty and mental effort. Each PD and ME were measured with six items (Krell, 2015) on a 5-point Likert scale (1="strongly disagree" to 5="strongly agree"). PD items assessed students' retrospective perception of the difficulty of a unit (e.g., "The contents of the tasks were complicated"); ME items requested students to reflect on their effort in completing the unit tasks (e.g., "I have done my best to complete the tasks"). Responses to items with an inverse wording were reverse coded (e.g., "The tasks were easy to solve."). Mean scores across items were determined per person and unit (PD: Cronbach's α between .85-.92 across units; ME: α between .80-.84 across units).

2.3 Procedure and Design

Taking about two hours, the test procedure was computer-based and took place in group settings with group sizes of up to 20 students, supervised by trained test administrators. The students gave informed consent for participation and were requested to fill in a questionnaire about demographic information (including graduation grades). Afterwards, they completed the MDC test. The participants randomly received three MDC units out of a pool of six units. In order to vary the assigned MDC units and their order in the test systematically, a balanced incomplete block design was applied, resulting in 60 testlets. The test administrators were instructed to point out to students that each participant receives different parts of the test material. All functionalities of the MDC units were introduced in a comprehensive video-based tutorial. Subsequent to each unit, participants were asked to reflect on the difficulty of the tasks within their unit (PD) and on their own efforts to solve it (ME). Afterwards they were given the opportunity to take a short break. The test ended after the participants completed their three assigned MDC units.

2.4 Data analysis

For the prediction of PD and ME, we tested linear mixed models (LMM); for testing the prediction of the probability to correctly solve an MDC item, a series of generalized linear mixed models were conducted (GLMMs). Accounting for a hierarchical data structure, these regression models allow for the inclusion of fixed and random effects in the prediction of a continuous dependent outcome (LMM) and a dichotomous dependent outcome (GLMM). Fixed effects refer to regression weights that are constant across persons or items; random effects reflect that the prediction varies across persons or items. For LMMs, the regression coefficients reflect the effects on PD or ME; for GLMMs, they reflect the effects on the probability to give a correct response to the MDC items in a logit metric.

Due to the planned missing data design, we have assumed that missing values occur randomly (MAR; e.g., van Buuren, 2012). The analyses were conducted in *R* (v.3.4.4; R Core Team, 2018) using the packages *LogFSM* (cf. Kroehne & Goldhammer, 2018), *TAM* (Robitzsch, Kiefer, & Wu, 2017), *psych* (Revelle, 2018) and *lme4* (Bates, Mächler, Bolker, & Walker, 2015). Note that the lmer() and glmer() functions of the *lme4* package apply a listwise deletion procedure to the data in long format. Baseline models were specified by modelling fixed effects of units and unit position as well as random effects of students. The GLMM baseline model additionally included the fixed effects of the unit characteristic and the random effects of items. The predictors were added to the baseline models for hypothesis testing (see appendix for overview). Since the documents within different units are not directly comparable, the fixed effects of units were always included to account for unit specificities that are not due to other investigated factors at unit level. Note that the fixed effects of units are not reported in the results. Continuous predictor variables were log-transformed as needed to take skew distributions into account and *z*-standardized.

3. Results

Upon inspecting students' PD and ME together with the unit difficulties visually (Fig.2), students showed an ability to be able to assess difficulty relatively well ($r(3)=.40$, n.s.). Their ME appeared to be comparable between units. PD and ME had a low correlation ($r=-.01$ to

.17 across units). Furthermore, PD was positively correlated with study level, indicating higher PD for students in a Master's program, processing times, and sourcing; higher ME scores were also associated with an enrollment in a Master's program, better graduation grades, longer processing times, a higher extent of corroboration, and note-taking (Tab.4). For the students who took notes, PD was negatively associated with notes reflecting comprehension monitoring, while ME was positively correlated with notes on source information (Tab.5).

«Include Tab.4, Tab.5»

3.1 Prediction of cognitive load

The baseline models for predicting cognitive load showed differences in students' PD (*SD* of the random intercepts=.41) and ME (*SD*=.53). Tab.6 shows the estimates of the predictors. PD was not explained by ME, but decreased for units administered later than the first unit (position 2: *B*=-0.21; position 3: *B*=-0.15). PD increased for units where students needed to process three instead of two texts (*B*=0.42; supporting H1) and longer documents (*B*=0.25; supporting H2), but showed no differences for units where students were required to write an essay at first (*B*=-0.01; rejecting H3). Concerning the individual and behavioral variables, Master's students reported a higher PD than Bachelor's students (*B*=0.19) and the PD was lower if students applied sourcing (*B*=-0.20). Graduation grades and other behavioral variables were unrelated. Furthermore, ME decreased by the total length of the documents (*B*=-0.09), was higher for Master's students than Bachelor's students (*B*=0.17), and increased by processing time (*B*=0.13). There were no relations between ME and other variables.

«Include here Tab.6»

3.2 Prediction of MDC

The baseline model for predicting the probability of a correct MDC item response showed that the random effects representing MDC varied between students (*SD*=.61) and the random effects representing easiness varied between items (*SD*=.93). Tab.7 shows the results of models that blockwise include the student characteristics (M1), behavioral variables (M2), aspects of cognitive load (M3), and finally, all predictors (M4). The results show almost no effects of the unit position (except for the second position in M1 and M2; rejecting H6). Although individual characteristics and other behaviors were controlled for, the probability of MDC item success was significantly explained in terms of the MDC-specific strategies of corroboration (M4: *B*=0.09) and sourcing (M4: *B*=0.23) (supporting H4a). The use of annotations, though, did not contribute significantly to the explanation of MDC item success (rejecting H4b). In M4, however, there was a significant negative effect of notes taken that referred to commonalities and differences between documents within units (M4: *B*=-0.28). As expected (H5), the probability of correctly solving an MDC item decreased by the PD (M4: *B*=-0.14) and increased by ME (M4: *B*=0.14), showing independent contributions of the predictors in explaining students' success in solving the MDC items.

«Include here Tab.7»

4. Discussion

This present study investigated the role of cognitive load in the comprehension of multiple documents. It aimed at examining the relationship between the characteristics of tasks, individuals and their behavior with the PD and ME that students experience as well as their success in MDC items over and above these characteristics. Our results showed that several variables at unit, person, and behavioral level explained the amount of PD and ME. Even after accounting for these variables, PD and ME contributed to predict success probability in MDC items slightly, but independently. This indicates that an increased PD reflecting the blocking of cognitive resources comes at the expense of comparing and integrating information from multiple documents, regardless of the ME invested into task processing, the empirical difficulties and characteristics of units, individuals, and behavior. In the next section, we reflect on the findings in detail.

4.1 Characteristics of tasks

The number of documents and the total length of documents positively predicted the amount of PD, which is in line with research on the split attention effect and the assumption of higher element interactivity in multiple documents (Ayres & Sweller, 2005). However, although students perceived the units that included more documents as more difficult, this had no effect on their comprehension. This might indicate that although element interactivity had increased, the students had still enough WM capacity available to process the given material. Therefore, the number of documents provided (2 vs. 3 documents) might not have been large enough to induce demands of information distribution sufficiently and produce cognitive overload in students. Alternatively, it is not the distribution of information across documents that takes up the processing resources required for MDC, but the overlap of information and the type of events which readers need to track (events related to time, space, protagonists/objects, intentionality, and causality; Therriault & Rinck, 2007). Redundancies in learning materials increase the cognitive load perceived (redundancy effect; Leppink & van den Heuvel, 2015). In the context of multiple documents, though, they might support instead of hinder comprehension by signaling consensus (i.e., “the sources agree on this point, therefore it must be true”). Taking the semantic overlap between documents and occurrence of events into account could generally provide insights into how students construct an integrated mental model of the situation or phenomena described in documents, and in particular which role WM plays in the construction of this model. However, the degree of redundancy between documents is challenging to define, as redundancy might be determined in terms of propositions on different micro or macro levels (e.g., referring to the same wording or the same message across documents). This should be addressed in future studies.

Concerning the prediction of ME, we found a negative effect of unit text length. At first sight, this might indicate a motivational effect; longer unit texts seem to have a daunting effect as readers report higher engagement when working with shorter units. Assuming that the processing of longer texts is associated with a higher investment of time and ME, and taking into account the positive effect of unit processing time on ME, this finding is indeed conclusive. Yet, it points to a basic problem of measuring ME. Although PD and ME were both assessed by self-ratings, the introspection of ME might be less objective compared to the introspection of PD. The self-assessment of one’s ME can depend on interactions between readers and tasks (e.g., “I spent more time on the task, therefore I invested more effort”) or the individual’s definition of engagement and successful task completion. Nevertheless self-assessments of cognitive load have proven to be valuable (Ayres, 2006), especially when combined with other procedures of assessing cognitive load, such as eye-tracking or different styles of self-rating (Klepsch et al., 2017; Korbach et al., 2018).

The third unit characteristic—the requirement to write an essay—did not show the expected effect of reducing PD in students. This result was observed while controlling for the number of documents and the total text length. It might not be overly stressed, though, as the sample contains only five units and unit topic might be confounded with the difficulty of the units. A systematic comparison of student groups who received the same unit with and without the essay requirement, for example, would make possible effects traceable.

4.2 Individual characteristics

With regard to the individual characteristics, graduation grades did not predict PD or ME, and they also did not affect the relationship of PD and ME to MDC, which shows that cognitive load is not an issue of lower general cognitive skills. However, the students' study level revealed an interesting result pattern. Taking into account graduation grades and behavioral variables, students enrolled in a Master's program did not perform significantly better on the MDC items than the Bachelor's students, but they reported a higher PD and ME. Although the study level cannot reflect how much actual experience students have in dealing with multiple documents (cf. Kobayashi, 2009), a possible explanation might still be that Bachelor's students in the first semester have lower standards of MDC than would be expected in the academic field or according to the normative assumptions of the documents model framework (Britt & Rouet, 2012). Hence, it might be troublesome for them to assess the difficulty of a MDC task accurately and adequately, and to adapt their behavior accordingly. The intercorrelations (Tab.2) add some evidence to this suggestion. Master's students showed more corroboration than Bachelor's students. Herein, PD and ME reveal conceptual overlap with procedural metacognition, i.e., the degree to which students monitor and control their own cognitive activity. Certain indicators of procedural metacognition (Nelson & Narens, 1990, for an overview) build on the accuracy of students assessing their own performance for a given task, which in turn is related to their judgements of task difficulty after performing the task. These retrospective judgments of performance (e.g., Händel, Artelt, & Weinert, 2013) are regarded as indicators of procedural metacognition because of their status as prerequisites for monitoring and controlling one's own cognitive activities.

4.3 Student behaviors

Only sourcing had an effect on PD, as units were perceived as easier when students engaged in sourcing activities. Sourcing is expected to be crucial for MDC, as it helps readers to interpret the document context and resolve conflicts that they might detect between documents (Bråten et al., 2018). In this regard, sourcing might even activate schemas that incorporate knowledge about specific text genres, and therefore provide an anticipatory framework for the subsequent encoding of a document (Paas, 1992; Wineburg, 1991). Corroboration had no predictive value for the aspects of cognitive load, but was predictive for MDC in a manner similar to sourcing, regardless of graduation grades. This is an interesting ancillary finding. It emphasizes that MDC is characterized by meaning-making activities that are not due to general cognitive skills. This supports the conclusion that mere confrontation of students with multiple documents does not necessarily lead to them achieving deep learning (Britt & Rouet, 2012). Accordingly, appropriate strategies for working with multiple documents need to be actively learned and provided via training so that they can be systematically executed (Britt & Aglinskas, 2002). It should be noted, however, that the behavioral variables used in this study may not represent the strategies studied sufficiently since the present operationalization does not necessarily reflect the effective implementation of a strategy. For example, attention to source information does not assure that students have the textual schemas available for encoding a document (Wineburg, 1991). A theoretically guided refinement of the behavioral indicators is necessary. Furthermore, it would be interesting to use behavioral data to identify episodes of serial and parallel processing. It is possible to process multiple documents serially at first and integrate the gained information across the documents afterwards. Some students might engage in this strategy, but it would create unnecessary cognitive load, as they have to determine correspondence between elements later on (redundancy effect).

Assuming that annotations serve to relieve the reader of cognitive resources by consolidating memory traces or offloading memory content to external locations (Hagen et al., 2014; Kobayashi, 2009; Moos, 2009), it was surprising to find no effects of using annotation functions, except for notes on intertextual comparisons of documents. The effect found even indicated that taking notes on intertextual comparisons is associated with reduced success in the MDC items. Since note-taking was not mandatory for students and they rarely took notes, we would rather interpret this result as a strategic attempt to overcome comprehension difficulties by students who struggled with the documents.

4.4 WM depletion

Last but not least, we investigated effects of the administered unit positions in order to account for possible effects of WM depletion (Chen et al., 2018). WM depletion occurs over a period of extensive mental effort and results in decreasing performance over time. We could not observe this in our data, but found that the PD decreased over time. Since the presentation and structure of the units did not change over time (Ayres, 2006), this might indicate a reduction of extraneous cognitive load, as students became familiar with the initially novel reading environment.

4.5 Limitations

There are limitations that need to be considered. Firstly, since the MDC test was aimed at measuring MDC as a generic competence of students independently of prior knowledge, the contents of the MDC units were mostly invented. It was assumed that students had little exposure with the exact content of the documents, and therefore did not possess prior knowledge in a narrow sense. However, this does not include prior knowledge in a broader sense (e.g., knowledge of related domains, or meta-knowledge about text genres), topic interest or even a feeling of knowing ("felt prior knowledge"). Effects that require a particular level of expertise should hardly occur due to the fictitious content provided in the MDC units. Nevertheless, this assumption should be tested in future studies by assessing, for example, topic interest or felt prior knowledge. Secondly, the primary objective of the study from which the investigated data originated was to develop an instrument to measure MDC. Substantially more items were constructed and presented to the students than were finally included in the analyses to answer the research questions. This might have led to a biased estimation of the investigated PD and ME effects, since the students rated the PD and ME items based on their experience with more MDC items than were included in the analyses. In terms of a replication study, corroborating the results using the final item set only is most desirable. Thirdly, PD and ME were assessed after students completed the MDC units. For future research, it would be of interest to gain insights about peaks in cognitive load (Paas et al., 2003) during unit processing, which would provide further insight in to the interdependence of MDC and cognitive load. However, self-report measures, in particular, could disrupt the comprehension process and change the response process itself. Alternatively, online measures (e.g., detecting facial expressions, eye tracking, log-file data) could be used to represent cognitive load, but evidence must be found that these measures can represent cognitive load sufficiently well (cf. Sweller et al., 2011). Finally, the results are based on an ad-hoc sample. In this sense, there was no purely random sampling, which means that the sample cannot claim to fully represent the student population. The results of our study can therefore not be generalized to students from the humanities and social sciences or other populations.

Nevertheless, this study provides interesting insights into the role of cognitive load for MDC. The PD that students experience when dealing with multiple documents can be explained in terms of MDC-specific requirements but cannot be compensated by student behavior. However, difficulties related to PD when creating a comprehensive mental representation of documents may be due to other factors that are not necessarily MDC-specific. A joint consideration of the relationship between cognitive load and MDC with students' skills in WM and reading comprehension will provide further insights.

References

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

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 135–146). New York: Cambridge University Press.

Ayres, Paul. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, *16*(5), 389–400. https://doi.org/10.1016/j.learninstruc.2006.09.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

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.), *Handbook of Discourse Processes*. New York, NY: Taylor & Francis.

Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, *20*(4), 485–522.

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). New York: Cambridge University Press.

Britt, M. A., & Sommer, J. (2004). Facilitating textual integration with macro-structure focusing tasks. *Reading Psychology*, *25*(4), 313–339. https://doi.org/10.1080/02702710490522658

Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2018). Extending Cognitive Load Theory to Incorporate Working Memory Resource Depletion: Evidence from the Spacing Effect. *Educational Psychology Review*, *30*(2), 483–501. https://doi.org/10.1007/s10648-017-9426-2

Chen, O., Kalyuga, S., & Sweller, J. (2017). The Expertise Reversal Effect is a Variant of the More General Element Interactivity Effect. *Educational Psychology Review*, *29*(2), 393–405. https://doi.org/10.1007/s10648-016-9359-1

Choi, H.-H., van Merriënboer, J. J. G., & Paas, F. (2014). Effects of the Physical Environment on Cognitive Load and Learning: Towards a New Model of Cognitive Load. *Educational Psychology Review*, *26*(2), 225–244. https://doi.org/10.1007/s10648-014-9262-6

DeStefano, D., & LeFevre, J.-A. (2007). Cognitive load in hypertext reading: A review. *Computers in Human Behavior*, *23*(3), 1616–1641. https://doi.org/10.1016/j.chb.2005.08.012

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The Time on Task Effect in Reading and Problem Solving Is Moderated by Task Difficulty and Skill: Insights From a Computer-Based Large-Scale Assessment. *Journal of Educational Psychology*, *106*(3), 608–626.

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*, 141–157. https://doi.org/10.1111/j.1467-9817.2012.01536.x

Händel, M., Artelt, C., & Weinert, S. (2013). Assessing metacognitive knowledge: Development and evaluation of a test instrument. *Journal of Educational Research Online*, *5*, 162–188.

Hutchins, E. (1995). How a Cockpit Remembers Its Speeds. *Cognitive Science*, *19*(3), 265–288.

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and Validation of Two Instruments Measuring Intrinsic, Extraneous, and Germane Cognitive Load. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.01997

Kobayashi, K. (2009). The influence of topic knowledge, external strategy use, and college experience on students’ comprehension of controversial texts. *Learning and Individual Differences*, *19*(1), 130–134. https://doi.org/10.1016/j.lindif.2008.06.001

Korbach, A., Brünken, R., & Park, B. (2018). Differentiating Different Types of Cognitive Load: a Comparison of Different Measures. *Educational Psychology Review*, *30*(2), 503–529. https://doi.org/10.1007/s10648-017-9404-8

Krell, M. (2015). Evaluating an instrument to measure mental load and mental effort using item response theory. *Science Education Review Letters, Research Letters 2015*, 1–6.

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*. https://doi.org/10.1007/s41237-018-0063-y

Landis, J. R., & Koch, G. (1977). The measurement of observer agreement for categorial data. *Biometrics*, *33*(1), 159–174.

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.

Leppink, J., & van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education*, *4*(3), 119–127. https://doi.org/10.1007/s40037-015-0192-x

Moos, D. C. (2009). Note-taking while learning hypermedia: Cognitive and motivational considerations. *Computers in Human Behavior*, *25*(5), 1120–1128. https://doi.org/10.1016/j.chb.2009.05.004

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. *The Psychology of Learning and Motivation*, *26*, 125–141.

Paas, F. (1992). Training Strategies for Attaining Transfer of Problem-Solving Skill in Statistics: A Cognitive-Load Approach. *Journal of Educational Psychology*, *84*(4), 429–434.

Paas, F., & Sweller, J. (2012). An Evolutionary Upgrade of Cognitive Load Theory: Using the Human Motor System and Collaboration to Support the Learning of Complex Cognitive Tasks. *Educational Psychology Review*, *24*(1), 27–45. https://doi.org/10.1007/s10648-011-9179-2

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, *38*(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension. *Scientific Studies of Reading*, *11*(4), 357–383.

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Revelle, W. (2018). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Evanston, Illinois: Northwestern University. Retrieved from https://CRAN.R-project.org/package=psych

Robitzsch, A., Kiefer, T., & Wu, M. (2017). *TAM: Test analysis modules*. Retrieved from https://CRAN.R-project.org/package=TAM

Rouet, J.-F., Ros, C., Goumi, A., Macedo-Rouet, M., & Dinet, J. (2011). The influence of surface and deep cues on primary and secondary school students' assessment of relevance in Web menus. *Learning and Instruction*, *21*(2), 205–219. https://doi.org/10.1016/j.learninstruc.2010.02.007

Scheiter, K., Gerjets, P., Vollmann, B., & Catrambone, R. (2009). The impact of learner characteristics on information utilization strategies, cognitive load experienced, and performance in hypermedia learning. *Learning and Instruction*, *19*(5), 387–401. https://doi.org/10.1016/j.learninstruc.2009.02.004

Schoor, C., Hahnel, C., Artelt, C., Reimann, D., Kröhne, U., & Goldhammer, F. (2018). 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, conditionally accepted*.

Sweller, J. (2010). Element Interactivity and Intrinsic, Extraneous, and Germane Cognitive Load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J., Ayres, P. L., & Kalyuga, S. (2011). *Cognitive load theory*. New York: Springer.

Therriault, D. J., & Rinck, M. (2007). Multidimensional Situation Models. In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain: inference and comprehension processes* (pp. 311–328). Mahwah, NJ: Erlbaum.

Trapmann, S., Hell, B., Weigand, S., & Schuler, H. (2007). Die Validität von Schulnoten zur Vorhersage des Studienerfolgs - eine Metaanalyse [The validity of school grades for predicting study success - a meta-analysis]. *Zeitschrift für Pädagogische Psychologie*, *21*(1), 11–27. https://doi.org/10.1024/1010-0652.21.1.11

van Buuren, S. (2012). *Flexible imputation of missing data*. Boca Raton, FL: CRC Press.

Weinstein, C. E., & Mayer, R. E. (1983). The teaching of learning strategies. *Innovation Abstracts*, *5*(32).

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source Evaluation, Comprehension, and Learning in Internet Science Inquiry

Tasks. *American Educational Research Journal*, *46*(4), 1060–1106. https://doi.org/10.3102/0002831209333183

Wineburg, 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.

Appendix

Overview on the specified models

Models	*R* syntax
LMMs: Prediction of PD and ME (Tab.6)	
Baseline model	PD/ME ~ 1 + unit + position + (1|student)
PD model (including all predictor variables)	PD ~ 1 + ME + individual variables + unit variables + behavioral variables + unit + position + (1|student)
ME model (including all predictor variables)	ME ~ 1 + PD + individual variables + unit variables + behavioral variables + unit + position + (1|student)
GLMMs: Prediction of MDC item success (Tab.7)	
Baseline model	item success ~ 1 + unit variables + unit + position + (1|student) + (1|items)
M1 (including individual variables as predictors)	item success ~ 1 + individual variables + unit variables + unit + position + (1|student) + (1|items)
M2 (including behavioral variables as predictors)	item success ~ 1 + behavioral variables + unit variables + unit + position + (1|student) + (1|items)
M3 (including cognitive load variables as predictors)	item success ~ 1 + PD/ME + unit variables + unit + position + (1|student) + (1|items)
M4 (including all predictor variables)	item success ~ 1 + PD/ME + individual variables + behavioral variables + unit variables + unit + position + (1|student) + (1|items)

Note. Fixed effects of units were always included but not reported in the results. The behavioral variables include the variables on note-taking.

Tables

Tab.1

Overview of the predictor variables

Variable	**Format in models**	**Value description**	***n***	***M* / RF**	***SD***
Variables on individual level					
study level	dichotomous	0 = Bachelor program 1 = Master program	297	0.31	-
graduation grades	*z*-standardized	small values indicate higher proficiency	305	2.20	0.62
Variables on unit level					
number of documents	dichotomous	0 = two documents within a unit 1 = three documents within a unit	5	0.60	-
unit text length	*z*-standardized	high values indicate higher word counts	5	1532.40	365.97
essay	dichotomous	0 = no essay required 1 = essay writing was requested	5	0.40	-
Variables on individual x unit level					
perceived difficulty	*z*-standardized	high values indicate that students perceived the tasks of a unit as difficult	774	2.83	0.83
mental effort	*z*-standardized	high values indicate that students invested high mental effort in unit processing	774	3.85	0.69
unit processing time	log-transformed and *z*-standardized	high values indicate longer processing times (in minutes)	774	25.76	7.31

Variable	Format in models	Value description	*n*	*M* / RF	*SD*
corroboration	added by 1, log-transformed and *z*-standardized	high values indicate more frequent switches between documents	774	14.72	9.32
sourcing	dichotomous	0 = not attended to all sources 1 = attended to all sources	774	0.39	-
comprehension monitoring	dichotomous	0 = does not contain notes of this category 1 = contains notes on comprehension monitoring	774	0.05	-
intratextual paraphrasing	dichotomous	0 = does not contain notes of this category 1 = contains notes with intratextual paraphrasing	774	0.33	-
intratextual overview	dichotomous	0 = does not contain notes of this category 1 = contains notes with intratextual overview	774	0.05	-
intertextual paraphrasing	dichotomous	0 = does not contain notes of this category 1 = contains notes with intertextual paraphrasing	774	0.04	-
intertextual comparison	dichotomous	0 = does not contain notes of this category 1 = contains notes with intertextual comparison	774	0.05	-
source information	dichotomous	0 = does not contain notes of this category 1 = contains notes on source information	774	0.07	-
no notes taken	dichotomous	0 = notes were taken 1 = notes were not taken	774	0.62	-

Notes. Column *n* shows the number of observations. *M* and *SD* are the mean and standard deviation of the unstandardized variables. RF is the relative frequency of category “1” of dichotomous variables.

Tab.2

Description of the MDC units

Unit	**Content**	**Framing of text sources**	**No. items**	**No. texts**	**Text length**	**Essay writing**	***n***
Catalano	biography of the fictitious Mafia boss Catalano	an extract from a database on criminal history and an economic newspaper article	11	2	1273	not required	154
2134	the arrival of aliens on Earth in 2134 as a future historical event	a lab report, a government report, and a political speech	11	3	1458	not required	156
Nothing	book reviews of the fictitious novel “Nothing”	two newspaper articles	13	2	1320	required	151
Animals	introductory textbooks on fictitious literary approaches for interpreting animals in novels	three excerpts from educational textbooks	17	3	2172	not required	153
Universe	physical-cosmological theories about the end of the universe	three newspaper articles	15	3	1439	required	160

Notes. Text length is the sum of the word counts of documents within a unit; column *n* shows how many students worked on a particular unit. The sixth unit was excluded from further investigations.

Tab.3

Categories for coding notes taken

Coding category	Description of the note content	Example
comprehension monitoring	evaluation of one’s own (non-)comprehension of the document (in the present study, only notes of non-comprehension occurred)	“positive or negative aspect?”, “??”
intratextual paraphrasing	exact repetition or (correct) paraphrase of the present text	“gravity of blame”
intratextual overview	heading for a document passage (not including formal headings, e.g., “introduction”)	“early years of Catalano”
intertextual paraphrasing	exact repetition or (correct) paraphrase of the content from another document within the unit	“Big Freeze”
intertextual comparison	comparison between at least two documents within the unit with respect to (correct) commonalities and differences	“here as well: forgive one-self”
source information	(correct) information about the source, either as exact repetition or as paraphrase of the source information	“-> source: textbook”

Tab.4

Pearson and tetrachoric correlations of the predictor variables (774 cases)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) perceived difficulty	-						
(2) mental effort	.07*	-					
(3) study level	.08*	.16***	-				
(4) graduation grades	-.02	-.09*	-.21***	-			
(5) unit processing time	.16***	.20***	.12***	-.05	-		
(6) corroboration	.02	.17***	.23***	-.16***	.37***	-	
(7) sourcing	.08*	.06	.05	-.13***	.02	.19***	-
(8) no note taking	-.02	-.08*	-.07	.01	-.28***	-.09*	-.04

Notes. For determining the correlations with of study level and graduation grades, these variables were disaggregated to the unit level. *p < .05, **p < .01, ***p < .001.

Tab.5

Pearson and tetrachoric correlations for note takers (subsample of 295 cases)

	(1)	(2)	(3)	(4)	(5)	(6)
(1) comprehension monitoring	-					
(2) intratextual paraphrasing	-.09	-				
(3) intratextual overview	-.22	.32*	-			
(4) intertextual paraphrasing	.31*	.04	-.39*	-		
(5) intertextual comparison	.30*	-.35**	-.14	.30*	-	
(6) source information	.04	-.19	.18	-.13	.03	-
Other predictor variables						
perceived difficulty	-.14*	.03	-.05	-.08	-.10	-.00
mental effort	.06	.02	.10	-.00	-.01	.16**
study level	.22	-.21	.15	.17	.34**	.41
graduation grades	.02	.04	-.03	-.07	-.09	-.13*
unit processing time	-.04	.14*	.02	.12*	.07	.03
corroboration	-.04	-.08	-.01	.21	.06	.11
sourcing	-.23*	-.02	-.04	.14	-.41***	.40***

Notes. For determining the correlations with of study level and graduation grades, these variables were disaggregated to the unit level. *p < .05, **p < .01, ***p < .001.

Tab.6

Prediction of perceived task difficulty (PD) and mental effort (ME)

Predictor	**PD**	**ME**
intercept	2.82 (0.14)***	3.72 (0.11)***
unit position 2	-0.21 (0.06)**	0.08 (0.04)
unit position 3	-0.15 (0.06)*	0.00 (0.04)
Aspects of cognitive load		
perceived difficulty	-	-0.03 (0.02)
mental effort	0.01 (0.03)	-
Unit characteristics		
number of documents	0.42 (0.10)***	0.08 (0.07)
unit text length	0.25 (0.06)***	-0.09 (0.04)*
essay	-0.01 (0.08)	-0.06 (0.06)
Individual characteristics		
study level	0.19 (0.07)*	0.17 (0.08)*
graduation grades	0.02 (0.03)	-0.04 (0.03)
Behavioral variables		
unit processing time	0.06 (0.04)	0.13 (0.03)***
corroboration	0.03 (0.04)	0.00 (0.03)
sourcing	-0.20 (0.08)**	0.08 (0.06)
Note taking behavior		
comprehension monitoring	-0.13 (0.13)	-0.01 (0.10)
intratextual paraphrasing	0.00 (0.13)	0.04 (0.10)
intratextual overview	-0.15 (0.13)	0.01 (0.09)
intertextual paraphrasing	-0.22 (0.14)	-0.07 (0.10)
intertextual comparison	-0.05 (0.13)	0.02 (0.09)
source information	0.10 (0.13)	0.06 (0.10)
no note taking	-0.02 (0.13)	0.05 (0.10)

Note. *p < .05, **p < .01, ***p < .001.

Tab.7

Predictions of the probability of a correct MDC item response

Predictor	**M1**	**M2**	**M3**	**M4**
intercept	0.25 (0.25)	0.11 (0.28)	0.29 (0.24)	0.16 (0.28)
unit position 2	0.13 (0.06)*	0.13 (0.06)*	0.06 (0.06)	0.10 (0.06)
unit position 3	0.03 (0.06)	0.05 (0.06)	0.00 (0.06)	0.05 (0.06)
Unit characteristics				
number of documents	-0.15 (0.42)	-0.28 (0.42)	-0.08 (0.42)	-0.21 (0.43)
unit text length	0.05 (0.24)	0.04 (0.24)	0.11 (0.24)	0.06 (0.25)
essay	0.72 (0.37)	0.75 (0.37)*	0.74 (0.37)*	0.70 (0.37)
Aspects of cognitive load				
perceived difficulty	-	-	-0.13 (0.03)***	-0.14 (0.03)***
mental effort	-	-	0.18 (0.03)***	0.14 (0.03)***
Individual characteristics				
study level	0.14 (0.09)	-	-	0.11 (0.08)
graduation grades	-0.31 (0.04)***	-	-	-0.28 (0.04)***
Behavioral variables				
processing time	-	0.08 (0.04)*	-	0.06 (0.04)
corroboration	-	0.12 (0.04)**	-	0.09 (0.04)*
sourcing	-	0.28 (0.08)***	-	0.23 (0.08)**
Note taking behavior				

Predictor	**M1**	**M2**	**M3**	**M4**
comprehension monitoring	-	-0.13 (0.13)	-	-0.16 (0.14)
intratextual paraphrasing	-	0.11 (0.15)	-	0.09 (0.14)
intratextual overview	-	0.09 (0.14)	-	0.05 (0.14)
intertextual paraphrasing	-	0.08 (0.15)	-	0.01 (0.15)
intertextual comparison	-	-0.19 (0.13)	-	-0.28 (0.13)*
source information	-	0.06 (0.14)	-	-0.07 (0.14)
no notes taken	-	0.18 (0.15)	-	0.12 (0.14)
% interindividual variance explained	29.60	15.78	12.34	47.16

Note. *$p < .05$, **$p < .01$, ***$p < .001$.

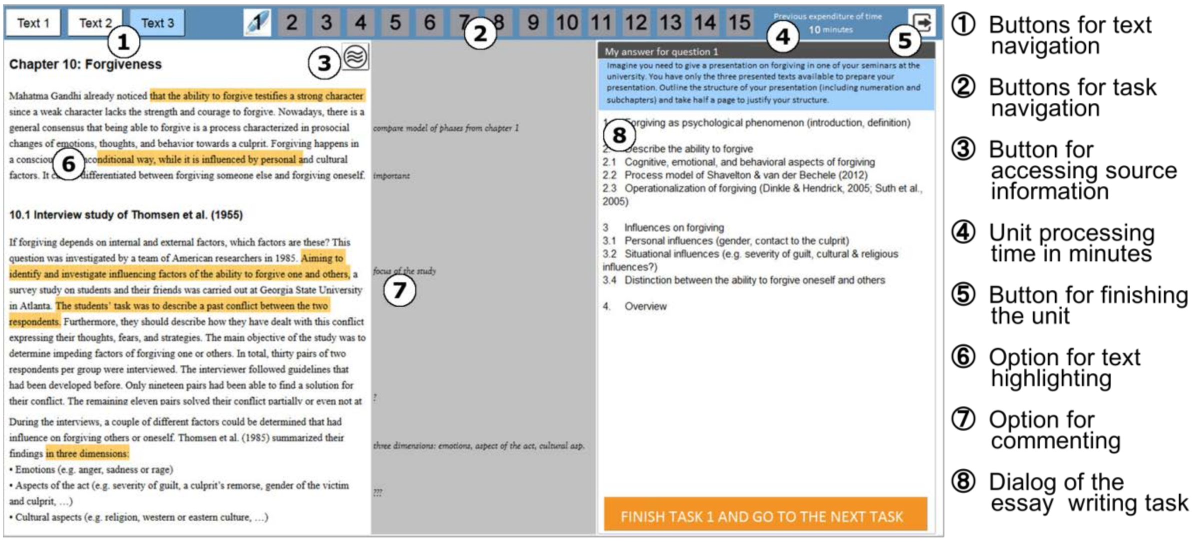

Fig.1. Example of an MDC unit.

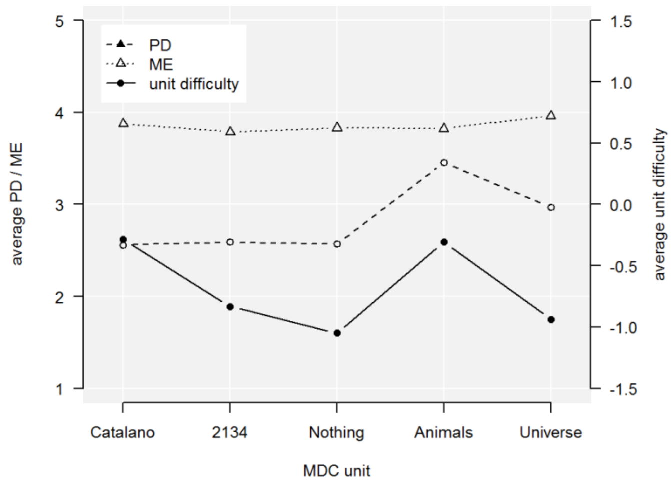

Fig.2. Average mean scores of perceived difficulty of tasks (PD) and mental effort (ME) across units and average difficulty of units. Unit difficulty was operationalized as the average of item difficulties within a unit, which were estimated based on the assumptions of the Rasch model.