Understanding Graphs: Modeling Processes, Prerequisites and Influencing Factors of Graphicacy

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Ulrich Ludewig

aus Steinfurt

Tübingen

2018

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	02.10.2018
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter:	Prof. Dr. Katharina Scheiter
2. Berichterstatter:	Prof. Dr. Augustin Kelava

Acknowledgments

This research was funded by the LEAD Graduate School & Research Network [GSC1028], a project of the Excellence Initiative of the German federal and state governments. Ulrich Ludewig is a doctoral student at the LEAD Graduate School & Research Network [GSC 1028]. Additionally, the Leibniz-Institut für Wissensmedien provided infrastructure and technical support.

My academic gratitude goes to my supervisors Prof. Dr. Katharina Scheiter, Prof. Dr. Augustin Kelava, and Prof. Dr. Korbinian Möller for introducing me into the world of science, and for helping me to develop my thinking and my writing. Additionally, to Prof. Dr. Wolfgang Schnotz, Prof. Dr. Erica De Vries and Prof. Dr. Neil Schwarz for providing the international cognitive visualization program which prepared me well for my dissertation.

Furthermore, I want to thank my colleagues and friends from AG 6 Multiple Representations and the LEAD Graduate School & Research Network for making work a joy, for many pleasant lunches and all of the off-work activities. I want to thank Cansu, Karo, Anne, Marie-Christine, Maria, Juliane and many more for proofreading my work.

I would also like to thank Selina Emhardt and Paul Feltes for the support during data collection; Mathias Rieckert, Adam Ayaita, and Dr. Michael Paul Grozs for many enlightening discussions about science; and Prof. Dr. Jürgen Heller, Dr. Stefan Huber, Dr. Simon Grund, Dr. Johann Jacoby, and Dr. Florian Wickelmaier for statistical support.

Moreover, I am very grateful for my family which supported me throughout my entire career.
Finally, I want to thank Pauline for being a supporting partner and for accepting the time constraints and the mobility which came along with this work. I would not have made it without your support.

Table of Content

Table of Content (2)

Table of Content (3)

Abbreviations

Abbreviation	**Written-out**
BLIM	Basic Local Independence Models
BNA	Basic Numerical Abilities
CDM	Cognitive Diagnostic Model
CKAO	Conceptual Knowledge about Arithmetic Operations
EMPPI	Explanatory Model of Person-Process Interaction
GLMER	Generalized Linear Mixed-Effect Regression
GC	Graph comprehension
PK	Prior Knowledge
IR	Internal Representation
IRT	Item Response Theory
KST	Knowledge Space Theory
LLTM	Linear Logistic Test Model
MIRT	Multi-Dimensional Item Response Theory
POMoG	Process-Oriented Model of Graphicacy
RC	Reading Comprehension
TC	Task Completion
TIMSS	Trends in International Mathematics and Science Study
ToT	Time-on-Task
TGT	Text-Graph Transitions

Chapter I. Introduction

'The evolution of the human mind is primarily the evolution of its means of expression', the French anthropologist André Leroi-Gourhan (1988, p. 262) concluded from his research on archeological artifacts. Leroi-Gourhan argued that the development of the human mind is closely related to the creation of symbol systems. Therefore, children spend their early years learning and practicing the use of linguistic and mathematical symbols. Nonetheless, as society changes, our means of expression continue to evolve. In the digital era, it has become essential to understand data quickly. The need to make data accessible to use led to the invention of data visualizations. In most cases, data is visualized as some type of graph (Friendly, 2008). Graphs represent our bank account balance, help us monitor our training progress, and display the weather forecast. Kosslyn (1989) defined graphs in his classic article 'Understanding Charts and Graphs' as representations with at least two scales with values associated with one another via a 'paired with' relation. Graphs represent greater quantities of the measured substance with more of some visual dimension (e.g., area, lines, diameter, angle, and color). Due to their abstract nature, understanding graphs is not trivial and can be viewed as a culture technique just like reading and operating with mathematical symbols, one that has to be taught, learned and practiced.

Today, the presence of graphs continues to increase (Friendly, 2008), and is expanding to many life sectors, workspaces and scientific domains due to the development of information technology. Therefore, understanding graphs is relevant across school subjects and contributes to students' reading, mathematical and scientific competences. Accordingly, many large-scale studies assess students' ability to understand graphs, for instance, as part of scientific literacy and mathematics in TIMSS (Baumert, Bos, & Watermann, 1998) or reading, mathematical and science literacy in PISA (OECD, 1999). These large-scale studies have shown that understanding graphs is a challenge for students. Strikingly, only 60 % of eight graders internationally can read a single value off of a line graph, and only 29 % can read an average off a graph (TIMSS 2011 Assessment, 2013). Both tasks represent the bare minimum of what students should be able to do. In light of these results, it is worthwhile to review the current state of research on how people understand graphs.

In fact, there is already a considerable amount of research on graphs across disciplines like literacy research, science and mathematical education, cognitive psychology, and medical and business decision-making. However, the research seems to be channeled into two distinct communities, the literacy and comprehension research communities. These research communities can be distinguished on the basis of their primary goals and research methods. On one side, literacy research aims at modeling individuals' competences to understand graphs. It investigates psychometric properties of tests, competence dimensions, item difficulty and factors influencing individuals' competence to understand graphs. On the other side, comprehension research aims at understanding how graph comprehension works. It investigates the effects of task, graph, content and individual characteristics and their interactions to explain the cognitive mechanisms underlying graph comprehension. Literacy research uses heterogeneous test items that simulate real-world problems, while comprehension research systematically manipulates task, graph, and content characteristics to enable inferences about cognitive mechanisms. Literacy research analyzes item responses from large heterogeneous samples to identify individual differences, whereas graph comprehension research analyzes various data sources like think-aloud protocols, eye movements, response times or response accuracy in homogenous, comparatively small samples to identify general graph comprehension processes (see Table 1).

Consequently, the strength of literacy research is its focus on realistic items and representative samples. Thus, literacy research addresses the question 'who is able to perform what real-world tasks?'; however, literacy research provides few insights into *how* individuals master these tasks. The strength of comprehension research lies in inferences about general graph comprehension processes. Therefore, comprehension research addresses the question of 'how do individuals achieve graph comprehension?'; however, the tasks in graph comprehension research are often artificial. In short, literacy research *describes* the ability to understand graphs, and graph comprehension *explains* how individuals understand graphs. Importantly, these research communities do not contradict, but rather complement each other. Therefore, this thesis attempts to combine literacy and graph comprehension research to understand 'who understands what graphs and how does it work?'.

The benefits of combining literacy and comprehension research are apparent. However, there are barriers to overcome before these two research communities can be integrated in a meaningful way. Integration requires terms, theoretical models and statistical modeling approaches that both communities can agree upon. Therefore, the first section of the introduction discusses the terminology used by both research communities. The second and third sections presents the current state of both research communities. The fourth section introduces the process-oriented model of graphicacy (POMoG), which integrates findings from both communities. The fifth section develops research questions that emerge from the POMoG before finally discussing modeling approaches that address these research questions. The introduction is followed by three studies (Chapters II, III, and IV) that address the research questions. Finally, the POMoG and the findings of the three studies will be discussed regarding their theoretical, methodological and practical implications.

Table 1. Comparison of the literacy and comprehension research communities concerning prototypical research goals, stimuli, study designs, data sources, and statistical modeling approaches.

	Literacy research	**Comprehension research**
Research goal	Describe individuals' ability to understand graphs: Test construction and competence modeling (dimensionality, item difficulty, and influencing factors)	Explain underlying graph comprehension processes: Effects of task, graph, content, individual characteristics and their interactions, as well as process measures
Stimuli	Test items that simulate real-world problems	Experimental conditions that manipulate task, graph, and content characteristics
Study designs	Large-scale assessment studies with representative samples	Experimental designs with process measures with homogeneous samples
Data sources	Item responses, test scores, and demographic variables	Response times, response accuracy, eye-tracking metrics, and think-aloud protocols
Statistical modeling approaches	Factor analysis (incl. IRT), Structural equation models, Regressions	Analysis of Variance, Analysis of Covariance, Qualitative analysis

I.1 Terminology across research disciplines

Many terms have been used to describe individuals' ability to understand graphs, due to the presence of graphs in different research disciplines. Only terms describing cognitive achievement dispositions that explicitly involve the comprehension of graphs[1] were considered here. The most frequent terms will be listed and discussed in order to identify commonalities and hone in on a conclusive definition.

The most frequently used terms to describe individuals' ability to understand graphically presented information are graphicacy (Åberg-Bengtsson & Ottosson, 2006; Lowrie, Diezmann, & Logan, 2012), graphical literacy (Galesic & Garcia-Retamero, 2011), graphing ability (Berg & Phillips, 1994; McKenzie & Padilla, 1986), graph sense (Friel, Curcio, & Bright, 2001), representational competence (Brenner, Herman, Ho, & Zimmer, 1999; Kohl & Finkelstein, 2005; Kozma & Russell, 1997; Stieff, Hegarty, & Deslongchamps, 2011), and representational fluency (Bieda & Nathan, 2009; Hill, Sharma, O'Byrne, & Airey, 2014). Table 2 provides an overview of constructs, definitions, target groups, representations, and authors.

The terms graphical literacy and graphicacy are used in the context of literacy research. Åberg-Bengtsson and Ottosson (2006) argue that "being '*graphicate*' is equal in status to being *literate* and *numerate*". They focus on individuals' ability to perform practical tasks which are likely to occur in everyday life. Galesic and Garcia-Retamero (2011) focused on graphs about health risks. Graphical literacy and graphicacy involve being able to extract information displayed in a graph and interpret it on the basis of a common knowledge context. In contrast, graph sense focuses more on the 'pure' extraction of information from graphs and a general understanding of coordinate systems and the apposed-position language. For Friel et al. (2001), graph sense can be understood as analogous to number sense (Sowder, 1992 as cited by Friel et al. 2001) and symbol sense (Fey, 1990, as cited by Friel et al. 2001).

However, graphical literacy, graphicacy, and graph sense focus on the ability to understand graphs independent of any specific content domain.

---

[1] *Constructs like 'visualizer' and 'visual literacy' as well as 'representational preference' are excluded because the first two refer to visual information more generally and the last refers to choice rather than performance.*

In contrast, representational competences, representational fluency and graphing ability focuses on individuals' ability to use graphs and other representations in conjunction with mathematical concepts or domain knowledge in the sciences. A critical aspect of these terms is that certain phenomena in mathematics and science can be displayed using different representational forms: for instance, population dynamics can be represented using time-persistent, time-implicate, and time-singular representations (Ainsworth & VanLabeke, 2004); molecular modeling as models, general equations, numerical equations, and graphs (Stieff et al., 2011); physics problems as words, graphs and equations (Hill et al., 2014). Kozma and Russell (1997) provide a frequently cited definition of representational competence (Brenner et al., 1999; Kohl & Finkelstein, 2005; Stieff et al., 2011). For them, representational competence is the set of skills for constructing, interpreting, transforming and coordinating domain-specific external representations for learning and problem-solving. Hill et al. (2014) use the term representational fluency and McKenzie and Padilla (1986) the term graphing abilities to describe constructs very similar to representational competence. Essentially, representational competence are its associated terms focus on understanding the specific content of a graph.

In sum, what all these terms have in common is that they describe individuals' ability to understand representations of data. However, the terms graphicacy, graphical literacy, and graph sense focus on individuals' ability to understand graphs independent of a specific content domain. In contrast, representational competence, representational fluency, and graphing ability focuses on individuals' ability to understand graphs and other representations in a specific content domain. Importantly, graphicacy is not content domain free; in principle, graphs representing real-world problems have a content domain. However, the content domain is not as critical for graphicacy as it is for representational competences. Furthermore, the terms related to representational competences involve a greater variety of operations, such as creating representations and transforming one representation into another, whereas as the terms around graphicacy instead focus on the interpretation of graphs.

The term graphicacy will be used in the following sections because this thesis examines individuals' ability to understand graphs independent of any specific content domain. Notably, Åberg-Bengtsson and Ottosson's (2006) argument that being '*graphicate*' is equal in status to being '*literate'* and '*numerate'* is rather misleading, because understanding graphs in real-world settings obviously relies on mathematical and reading skills in many cases. Therefore, graphicacy is instead a subset of existing literacy constructs, such as mathematical literacy and reading literacy, rather than an independent and novel construct. Nonetheless, it has already been pointed out that focusing on this sub-set is worthwhile. Subsequently, graphicacy is defined as the ability to understand graphs independently of specific content domains. A graphicate individual is able to understand the graphical language and content of a graph.

Table 2. Frequent terms describing the ability to understand graphs across research disciplines, with associated focus, relevant groups, representations and authors.

Construct	Focus	Group	Representation	Example authors
Graphical literacy	Real-life problem solving and decision making	Adults & secondary education	Bar, pie and line graphs	Galesic, & Garcia-Retamero (2011)
Graph sense	Mathematical problem solving	Primary & secondary education	Bar, pie, line graphs, scatter plots	Friel et al. (2001)
Graphicacy	Mathematical problem solving	Secondary education	Bar, line and picto graphs	Lowrie et al. (2012) Åberg-Bengtsson & Ottosson (2006)
Graphing ability	Utilizing graphs in science	Higher education	Graphs, tables, descriptions	McKenzie & Padilla (1986)
Representational competence	Working with domain-specific representations	Secondary education	Notations, graphs, equations, functions, tables	Kozma & Russell (1997); Stieff et al. (2011); Brenner et al. (1999)
Representational fluency	Problem-solving in the sciences	Higher education		Hill et al. (2014); Bieda & Nathan (2009)

I.2 Modeling graphicacy as a competence

This section addresses literacy research on graphs. Literacy research aims to model graphicacy as a competence. Therefore, in this section, the concept of 'competence', competence frameworks and competence modeling are introduced before reviewing the literature on graphicacy.

Competence is a popular concept in the cognitive, social and educational sciences. The keyword competence is contained in 14% of all publications in educational research and also occurs in many other scientific disciplines such as psychology, management, science education, health, economics, sociology and business[2]. As a result, the term 'competence' has many definitions and meanings. This thesis refers to competences in a pragmatic-functionalist sense, and specifically as a 'cognitive achievement disposition, that is functional with regard to situations and demands in a certain domain' (Prenzel, Gogolin, & Krüger, 2008, p. 14). Because competencies are situation- and domain-specific, the first step in modeling a competence is to develop a framework defining the domain of that competence. For instance, the competence framework for the Trends in International Mathematics and Science Study (TIMSS) defines *mathematics* as a combination of content (e.g., numbers, data, geometry) and cognitive domains (i.e., knowing, applying, and reasoning). The mathematics test in TIMSS contains items that meet the criteria of the competence framework. Items are sub-tasks of a test which test takers can answer correctly or incorrectly. A competence model - in the sense it is used here - is a statistical model that describes a competence based on test takers' responses to the items. Ideally, the competence model will be a parsimonious description of all responses. A competence model can be used to evaluate the validity and reliability of the associated test, as well as dimensionality and factors related to item difficulty. Of these, dimensionality and item difficulty are the most interesting from a substantive research perspective.

Statistically speaking, a dimension in a competence model is the conceptual equivalent of a 'factor' in a factor analysis. A dimension or factor is an unobserved latent variable that explains the joint variation in multiple item responses. Dimensionality determines what degree of differentiation is needed to describe the competence. In the case of graphicacy, it is only appropriate to say 'people can be more or less proficient in understanding graphs' if the competence model is one dimensional. Multiple dimensions imply that people have strengths and weaknesses. Therefore, it

[2] *Status 10/07/2018 Web of Science*

is necessary to specify people's competence on each competence dimension: for instance, 'people can be more or less proficient at reading points and interpreting trends'. In a multi-dimensional competence, reading points and interpreting trends would be two different aspects, meaning that being good at reading points does not mean one is necessarily good at interpreting trends as well.

Item difficulty is the probability of achieving a correct response at a given level of competence. Item difficulty can be influenced by item characteristics, e.g., graph complexity or task demands. Knowing which characteristics make items difficult is important for interpreting test results. Finally, the last aspect competence modeling investigates is what individual characteristics influence the competence. These correlations with other constructs are important for construct validity. For instance, graphicacy should correlate with reading comprehension; however, a very high correlation could imply that it is not necessary to assess graphicacy in addition to reading comprehension.

In sum, competence modeling has two crucial steps: first, developing a competence framework that defines the competence, and second, finding a parsimonious competence model that describes the dimensionality and factors related to item difficulty. Notably, the competence model is often expected to mirror the competence framework. However, the congruence between framework and model can be studied empirically. In the existing literature, three studies actually investigate the congruence between framework and model (Lachmayer, 2008; Nitsch & Bruder, 2014; Nitz, Ainsworth, Nerdel, & Prechtl, 2014; Ullrich et al., 2012), three studies focus on the framework and use a one dimensional competence model by default (Curcio, 1987; Hill et al., 2014; McKenzie & Padilla, 1986; Lai et al., 2016), and two studies specify competence models on the basis of test data without previously defining a competence framework (Åberg-Bengtsson, 1999; Åberg-Bengtsson & Ottosson, 2006). Therefore, competence frameworks and competence models for graphicacy are reviewed separately in the following sections.

I.2.1 Competence frameworks of graphicacy

This section reviews the graphicacy frameworks proposed by various researchers. Graphicacy frameworks define which task demands a graphicate person should be able to master. Most frameworks focus on graph interpretation; however, some additionally include graph construction and the integration of or translations to other representations and graphs. Furthermore, the frameworks have different content domains, such as science or decision making, depending on the research discipline from which they stem.

Lachmayer (2008) proposed a comprehensive framework for graph competences in the context of biology for ninth and tenth graders. The authors' framework has three main components: information extraction from graphs, graph construction, and integration. Each component has subtasks. Information extraction includes identification and graph reading. The first component, identification, refers to the ability to understand how the graph displays data and does not require the interpretation of the data itself. Graph reading is the ability to interpret the data itself. Graph reading includes four complexity levels: reading single values, reading a comparison or trend, reading multiple comparisons or trends, and reading beyond the data, which includes extrapolation and making predictions based on the data. The second component is graph construction and includes the construction of a graph frame and entering data. Construction of a graph frame is the ability to construct a graph structure that matches the variables to be depicted. Entering data encompasses three complexity levels, namely the ability to enter single data points, trends and multiple trends into a graph frame. The third component, integration, requires integrating information from graphs and other information sources, such as text and equations. Ullrich et al. (2012) focused specifically on the integration component for fifth to eighth graders in biology and geography. Similarly to Lachmayer's information extraction component, Ullrich et al. (2012) included three complexity levels: mapping single data points, mapping simple relations, and mapping complex relations.

Lai et al. (2016) proposed a framework including the three components of graph comprehension, critique and construction in science for fifth to eighth graders. The first of these, graph comprehension, includes locating the coordinates of a point and identifying relative highs; interpreting general relationships; describing shapes, trends, and noise in the depicted graphs; and integrating graphs with science ideas. Second, graph critique requires students to give alternative interpretations of graphs. The third element refers to the construction of graphs from tables and science concepts.

Nitz et al. (2014) proposed a framework for representational competence in biology. Their framework includes describing scientific concepts; generating and selecting a representation; identifying, describing and analysing features of representations; and making connections across different representations and explaining the relationships between them (Kozma & Russell, 2005).

McKenzie and Padilla (1986) proposed a framework for the Test of Graphing in Science (TOGS) for seventh to 12th graders. They defined eight different cognitive operations involving line graphs and scatterplots. These task demands are as follows: 1. selecting an appropriate scale and set of axes (given a description of an investigation), 2. Selecting a graph to display data (given a description of an investigation), 3. Selecting corresponding values for Y (or X) (given the opposing value), 4. Selecting an appropriate description of the relationship (given a graphed relationship), 5. Identifying graphs with appropriately assigned variables (given a series of graphs), 6. Identifying trends displayed in a set of data (given a graph), 7. Locating corresponding points on a graph (given coordinates), 8. Generalizing the interrelationship between two graphs (given two graphs)

Nitsch & Bruder (2014) propose a framework for students' ability to translate between representations of functions, including graphs, for ninth to tenth graders. The authors' framework includes translating between four types of representations (i.e. graph, numerical table, algebraic equation, and situation description) in both directions. However, they included on three translations that involve graphs: translating between graphs and algebraic equations, between graphs and numerical tables, and between graphs and situational descriptions.

Leinhardt, Zaslavsky, and Stein (1990) proposed a framework for the comprehension of functions with graphs within mathematics education by defining common misconceptions: confusing the slope and the height, confusing an interval and a point, considering a graph to be a picture, and conceiving of a graph as made up of discrete points.

However, Friel et al. (2001) focused more on the development of the skills needed to interpret graphs as mathematical constructs. They distinguished between three main components of graphicacy, progressing from local to global features of a graph: (a) reading information directly from a graph; (b) manipulating the information read from a graph by making comparisons and performing computations; and (c) generalizing, predicting, or identifying trends by relating the information in the graph to the situational context.

For Curcio (1987), graphicacy involves line, bar, and circle graphs, as well as pictographs. Students need to be able to read the data, title, and axis labels, as well as use their mathematical competence to make extensions, predictions, and inferences. Hill et al. (2014) proposed a similar framework for college physics education. The physics problems they discuss include questions that involve various combinations of graphs, verbal descriptions and equations.

I.2.2 Competence models of graphicacy

A competence model describes the empirical structure of a competence. This section describes the dimensionality and item difficulty of competence models of graphicacy in relation to their respective competence frameworks. Competence frameworks define multiple components of graphicacy. However, competence models examine which components actually need to be distinguished. For instance, Lachmayer (2008) compare the fit of eight hypothetical competence models to test results for N = 289 students (G9: n = 134, G10: n = 155). The eight models varied in granularity. The model that distinguished between five components - identification, graph reading, constructing a graph frame, entering data, and integration - had the best fit. This five-dimensional model had slightly better fit than a three-dimensional competence model distinguishing between information extraction, information construction, and integration. This implies that identification, graph reading, constructing a graph frame, entering data, and integration all require different competencies.

Ullrich et al. (2012) compared a one-dimensional model of integration and a three-dimensional model distinguishing between mapping single data points, mapping simple relations, and mapping complex relations using test results from N = 1060 students from grades five to eight. The one-dimensional competence model had the best fit. This indicated that the three types of mapping complexity require the same 'integration' competence.

Nitsch & Bruder (2014) found on the basis of data from 645 ninth and tenth graders that each type of translation between different representations of functions requires a specific competence (e.g., graph to description, table to graph). They compared the fit of a one-dimensional competence model for 'general comprehension of functions', a three-dimensional model for 'comprehension of representation forms' (e.g., graph, table and description), and a five-dimensional model of 'comprehension of translations' (e.g., graph to description). They found that students vary not only in their general comprehension of functions and their 'comprehension of forms of representation', but also in their ability to perform different types of translations. This shows that competences are not just specific to forms of representation, they are even specific to certain cognitive operations.

Åberg-Bengtsson (1999) analyzed a subset of the Swedish Scholastic Aptitude Test (SweSAT) involving the reading and interpretation of quantitative data. Åberg-Bengtsson performed a confirmatory factor analysis on 20 DTM items using data from 14,463 students from the same cohort. The result was a three-factor model with general, mathematical-quantitative and complexity competence competents. Åberg-Bengtsson and Ottosson followed-up on this work in 2006 by trying to distinguish among different factors underlying students' performance on a self-designed instrument for measuring graphicacy in a sample of 363 students between 15-16 years old from five schools. Again, the authors found three dimensions: a general graphicacy, an end-of-test and a narrative component. The narrative competent encompassed items with open-ended questions. Interestingly, the type of graph displayed (e.g., pie chart, line graph) and the complexity of the item (e.g., reading a single value, multiple choice) were not separable components. Åberg-Bengtsson's work shows that graphicacy tasks involving mathematical operations require a competence that needs to be distinguished from 'general' graphicacy.

Item difficulty. Item difficulty was not a primary goal in most of the studies reviewed here. However, Lachmayer (2008) found that complexity level (reading a single value, reading a comparison or a trend, reading multiple comparisons, reading beyond the data) explained item difficulty for the graph reading component. Furthermore, integration items were more difficult on average than identification and construction items. Ullrich et al.'s (2012) findings provide further support for the influence of complexity level in the context of integration. Earlier, it was emphasized how important it is to know which item characteristics make items difficult; however, too little work has been conducted in this area to give a comprehensive review.

I.2.3 Influencing factors

Content (domain) knowledge. Roth and Bowen (2001) argued that familiarity with the content of graphs is important for graphical competencies. Nitz et al. (2014) found that content domain knowledge and representational competence have a medium to large correlation. The authors argue that representational competence and content knowledge are 'interactively' related, but still empirically distinguishable. In line with Kozma and Russell (2005), Nitz et al. (2014) found that content knowledge improves faster than representational competence over the course of a teaching unit in a sample of N = 1253 students.

Reading and mathematical achievement. Curcio (1987) investigated how reading achievement (measured with the SRA Reading inventory), mathematics achievement (measured with the SRA mathematical inventory) and prior knowledge of the topic, the mathematical content and the graphical format predict graph comprehension with a sample of fourth (n = 204) and seventh graders (n = 185). The prior knowledge inventory covered items about the topic, mathematical content and the graphical form. Even when controlling for reading and mathematical achievement, the different prior knowledge components still correlated with graph comprehension. Prior knowledge of mathematical content had the highest correlation (r = .34 in grade seven and r = .38 in grade four). Interestingly, prior knowledge of the graphical format was only correlated with graph comprehension among the 4th graders. This may suggest that prior knowledge of the graphical format only influences graph comprehension when graphs are quite new in the curriculum. A regression analysis with all of the aforementioned variables revealed that mathematical and reading achievement are the best predictors of graph comprehension, whereas knowledge of the graphical format contributes little to prediction.

Åberg-Bengtsson and Ottosson (2006) found medium latent correlations between graphicacy and general academic achievement ($r = .48$), mathematical achievement ($r = .58$), and language achievement ($r = .47$). Galesic and Garcia-Retamero (2011) found correlations between graphical literacy and numerical literacy of $r = .47$ in a German sample and $r = .50$ in the US sample. These correlations show that the ability to understand graphs is related but not identical to general, mathematical and language achievement.

General cognitive abilities. Additionally, there is evidence that quantitative graph reading is associated with general cognitive abilities. For instance, Berg and Phillips (1994) found that performance in quantitative graph reading was positively associated with logical thinking and proportional reasoning in a sample of seventh, ninth, and eleventh graders. Moreover, Padilla, McKenzie and Shaw (1986) observed that interpreting line graphs was associated with abstract-reasoning abilities in a sample of 119 seventh, ninth, and eleventh grades.

I.2.4 Summary

A number of authors have proposed frameworks for graphicacy. These frameworks frequently distinguish between graph interpretation, graph construction and integration (Lachmayer, 2008; Lai et al., 2016; Leinhardt et al., 1990; Nitsch & Bruder, 2014). Within graph interpretation, researchers distinguish between different levels of complexity: reading single points, trends and complex relations (Lachmayer, 2008; Lai et al., 2016; Ullrich et al., 2012), as well as making generalizations and predictions (Lachmayer, 2008; Leinhardt et al., 1990). Additionally, there is a distinction between researchers focusing on individuals' abilities to understand the graph itself (von Kotzebue & Nerdel, 2015; McKenzie & Padilla, 1986) vs. what the graph represents (scientific phenomena: Lai et al., 2016, risk in medical contexts: Galesic, & Garcia-Retamero, 2011; biological phenomena: Lachmayer, 2008; functions: Leinhardt et al., 1990; phenomena in geography, biology and math: Åberg-Bengtsson & Ottosson, 2006).

Studies have shown that the distinction between interpretations, construction, and integration is in fact supported by competence dimensions from test results (Lachmayer, 2008; Nitsch & Bruder, 2014). This implies that the interpretation and construction of graphs and the integration of information from graphs and other sources require different competencies. In contrast, item characteristics such as graph type (i.e., bar and line graphs; Åberg-Bengtsson & Ottosson, 2006) and the complexity level of graph interpretation are not distinct competence dimensions (Lachmayer, 2008; Ullrich et al., 2012). Therefore, solving graphicacy tasks related to different types of graphs and making interpretations at different complexity levels require the same competence, even though more complex interpretation tasks are more difficult (Lachmayer, 2008).

Furthermore, graphicacy is related to domain knowledge (Nitz et al., 2014), reading and mathematical performance (Åberg-Bengtsson & Ottosson, 2006; Curcio, 1987; Galesic & Garcia-Retamero, 2011), and general cognitive abilities (e.g., Berg & Phillips, 1994). However, the pure association between graphicacy and content knowledge and reading and mathematical performance is difficult to ascertain, because the graphicacy tests used in previous studies have required test takers to apply content knowledge, read texts, conduct calculations, and/or apply mathematical concepts. It is unclear whether the influence of reading and mathematical performance on graphicacy is attributable to common cognitive mechanisms or simply to the reading and calculations required in graphicacy tests. Interestingly, Nitz et al. (2014) fond that representational competences develop as a result of domain knowledge rather than vice versa. This indicates that graphicacy in science is the end stage of scientific literacy rather than a prerequisite.

In sum, graphicacy frameworks predominantly agree on the task demands for graphicacy, and competence dimensions can be distinguished on the basis of operations, i.e. interpretation, construction, and integration. Therefore, it may be advisable to focus on the most researched task demands, namely interpreting graphs and integrating graphs and other information sources. Furthermore, whether one focuses on understanding the graph itself or the content of the graph makes a difference. The research reviewed in this section defines graphicacy and examines its components. However, it does not explain how graph comprehension works, which instead falls under the purview of graph comprehension research. Therefore, the next section will review research on graph comprehension to explore underlying graph comprehension processes.

I.3 Graph comprehension as a process

Graphicacy research aims to describe individuals' ability to understand graphs via construction tests and by modeling graphicacy as a competence. Now research from the comprehension research community will be discussed. Graph comprehension research aims at explaining underlying comprehension processes by investigating the effect of task, graph, content, and individual characteristics and their interactions, as well as process measures. The two research communities are connected because graphicacy is the ability to understand graphs. In other words, one could say that graphicacy is the individual disposition for graph comprehension. The core assumption of graph comprehension is that comprehension processes take place in order to construct internal representations. These comprehension processes are influenced by individual characteristics (e.g., domain content knowledge or visual-spatial abilities) and stimulus characteristics (e.g., graph, task, and content). In the following sections, two models on graph comprehension will be reviewed: the Model of Display Comprehension (Shah, Freedman, & Vekiri, 2005) and the Componential Model of Human Interaction with Graphs (Gillan & Lewis, 1994; Gillan, 2009). The Model of Display Comprehension and Model of Human Interaction with Graphs address graph comprehension with different levels of granularity. Shah et al. (2005) focus on conceptual comprehension processes involving domain knowledge and graph schemata, while Gillan and Lewis (1994) focus on the more perceptual side of graph comprehension, such as visual, visual imagery and mental processes. The models do not contradict, but complement each other. Therefore, the two models will be described individually first and then combined.

I.3.1 The Model of Display Comprehension (Shah, Freedman, & Vekiri, 2005)

The Model of Display Comprehension (Shah et al., 2005) is an adaptation of Pinker (1990). Shah et al. (2005) use the term 'display comprehension'; however, since the authors predominantly refer to graphs, it is considered a graph comprehension model. In Pinker's (1990) original model, he considered visual features of the display, gestalt processes, and the graph schema as factors that allow the user to extract conceptual information from a graph. This model of graph comprehension can be summarized in seven processing steps: (1) the user has a goal of extracting a specific piece of information; (2) the user looks at the graph, activating graph schema and gestalt processes; (3) the user encodes salient features of the graph on the basis of gestalt principles; (4) the user now knows which cognitive strategies to use; (5) the user then extracts goal-directed visual chunks; (6)

the user may compare these visual chunks; and (7) finally, the user extracts the relevant information to achieve the goal.

Freedman and Shah (2002) further developed the model by differentiating the influence of prior knowledge. In their model, comprehension of graphs is influenced by knowledge and graph characteristics. The influence of knowledge on comprehension is considered a ‘top-down’ process, while the influence of graph characteristics is considered ‘bottom-up’. The model distinguishes between two types of knowledge, display (or more specifically graph) schemata and content knowledge, and two types of internal representations, the internal representation of the display and the internal representation of the referent (see Figure 1). Graph schemata are specific to a certain graph type and can be seen as an active, interrelated knowledge structure of what this graph type is for and how it is interpreted in general (Pinker, 1990). A graph schemata translates the information found in the graph into conceptual information and directs the search for relevant pieces of information. Content knowledge is knowledge about what the graph represents and helps the viewer focus on relevant data and distinguish signals from noise. Content knowledge allows the view to draw inferences and learn from the graph.

Shah et al. (2005) described multiple processing steps for graph comprehension. First, the viewer encodes the visual features of the external display by focusing their attention on visual features; this visual attention is guided by graph schemata and domain knowledge. The viewer then uses graph schemata to build an internal representation of the display. The viewer then makes inferences using long-term knowledge and constructs an internal representation of the referent.

Shah, Mayer, and Hegarty (1999) assume that these steps take place in an iterative construction-integration process of graph comprehension, analogous to Kintsch’s (1988) construction-integration model for text. As in Kintsch’s theory, Shah et al. (1999) assume visual chunking. Visual chunks can be automatically linked to a quantitative fact. This linking process is more likely to be successful when the quantitative information is directly available in the visual chunk. There can be two reasons why the quantitative information is not available in the visual chunk. First, the viewer might lack the knowledge of graphical conventions to link the visual chunk with the quantitative fact. Second, the visual chunk may not be directly linked to the quantitative fact. For instance, the viewer may need to compare the mean height of multiple grouped bars. According to Shah et al. (1999), viewers are less likely to be successful when a complex inferential process is necessary to extract information. Because of visual chunking, Freedman and Shah (2002) stated that ‘graphical displays are most useful when they make quantitative information perceptually obvious’ (p.22). In other words, the way information is graphically displayed has to fit to the task. A good graph-task fit facilitates information extraction. However, in many cases graphs cannot be designed in a way that makes the quantitative information perceptually obvious, especially when the data and tasks are complex.

Trickett and Trafton (2006) investigate how experts work on complex data visualizations. They analyzed experts’ think-aloud protocols of their work, finding that experts use spatial transformations more frequently than any other cognitive strategy to conduct their work. The experts performed spatial transformations whenever information was not directly accessible (Trafton et al., 2000). Spatial transformations are any manipulation of mental images, such as adding or deleting features, mentally rotating features, mentally moving features, animating a static image, making comparisons between features, and other mental operations that transform the spatial array of the graph from one state to another. Additionally, they found that experts use far more spatial processing than novices (Trafton et al., 2000).

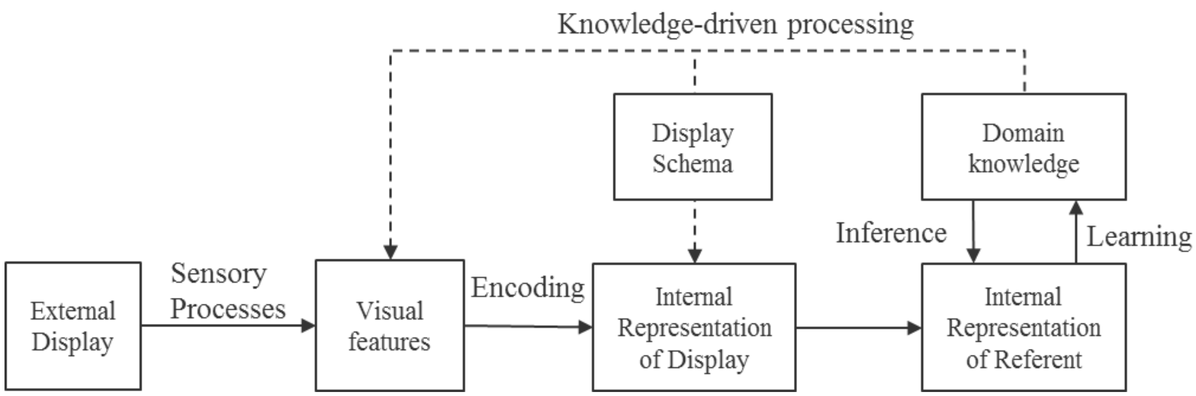


Figure 1. Model of comprehension of visual displays by Shah et al., 2005, indicating the interaction of bottom-up processes (solid arrows) and top-down processes (dashed arrows) in the construction of a mental model of the referent.

I.3.2 The Componential Model of Human Interaction with Graphs (Gillan & Lewis, 1994)

The Componential Model of Human Interaction with Graphs (Gillan & Lewis, 1994; Gillan, 2009) focuses more in detail on the manipulation, comparison and computation of information represented in graphs. Gillan and Lewis (1994) describe graph comprehension as a sequence of process components, such as searching for indicators, encoding indicators, performing arithmetic operations on the values, making spatial comparisons among indicators, and responding to the question. The model begins with a goal defined in relation to the viewer’s task (e.g., ‘compare the mean points of A and B'). The combination of the task, graph, and viewer's knowledge determines the sequence of the processing components that are then applied to meet the task demands. The model distinguishes three different types of process components involved in graph reading: arithmetic, visual and visual imagery processes. Arithmetic processes involve operations such as the addition, subtraction, division, and multiplication of numbers. Arithmetic operations are performed mentally. Visual processes encompass visual search, spatial comparisons of height and length, encoding values, and determining spatial differences. In contrast to visual imagery process, visual processes are associated with actual viewing behavior. Visual imagery processes are manipulations of a visual mental image, similar to Trickett and Trafton's (2006) spatial transformations (e.g., moving an image, rotation, competition, differentiation and selection of anchors).

These processing components are involved in solving graph reading tasks; however, the exact sequence of the processing components depends on the task, the graph and the reader's knowledge. For instance, when the task is to find the mean of two bars in a bar graph with the labels A and B, a viewer typically first conducts a visual search for label A, then scans over to the Y-axis to encode the value of the bar labeled A. The viewer repeats these two steps for label B, retaining the encoded values in working memory. Then, after encoding both values, the viewer adds the values to determine the sum and divides by two to determine the mean. However, a viewer proficient in visual imagery processing might apply a different sequence of processing components. The alternative processing steps are conducting a visual search to estimate the midpoint of the tops of the bars, then scanning over to the y-axis to encode that value, which is the mean of the values.

Gillan's Componential Model of Human Interaction with Graphs is specifically tailored to graph comprehension tasks that involve arithmetic operations. However, the model also identifies processing components that are most likely involved in all graph comprehension tasks. Moreover, the model demonstrates that the exact sequence of processing components depends on the exact task, the graph, and individual characteristics. Therefore, it proposes a rationale for explaining individual differences in graph comprehension.

I.3.3 Summary

The preceding sections described the Model of Display Comprehension by Shah et al. (2005) and the Componential Model of Human Interaction with Graphs by Gillan and Lewis (1994). Both models describe process components of graph comprehension. Shah et al. (2005) focus of higher-order comprehension processes, such as the application of graph schemata (i.e., knowledge of the graph conventions) and domain knowledge and making inferences. They distinguish between viewers' internal representation of the display and internal representation of the referent. Graph schemata influence the construction of the internal representation of the graph, and domain knowledge influences the internal representation of the referent. Gillan and Lewis (1994) focus on the fine-grained process components involved in graph comprehension. They distinguish between visual processing, visual imagery processes and mental processes. Visual processing components such as visual search, encoding, mapping, and comparing are associated with viewing behavior. Visual imagery processes are manipulations of a mental image, such as moving elements mentally to compare them to others. Visual imagery processes need be applied whenever information is not directly displayed in a graph. Gillan and Lewis (1994) originally saw these processing components as aligned in a sequence; however, for more complex graph comprehension tasks these processing components could also be part of a recurring construction-integration cycle. Gillan (2009) pointed out that which processing components are applied and in what sequence depends on a combination of task, graph and most importantly individual characteristics. Gillan does not explicitly explain these individual characteristics; however, Shah et al. (2005) addressed individual characteristics by highlighting graph schemata and domain knowledge, both of which can change which processing components individuals apply. This is where the two models connect. Visual processes are applied to build the internal representation of the graph, whereas the visual and visual imagery processes are applied to build the internal representation of the referent.

In any event, these models draw a detailed picture of how graph comprehension works. Now, how are graph comprehension processes related to frameworks and models of graphicacy? This question will be addressed in the following section, which integrates graphicacy and graph comprehension research to develop a process-oriented model of graphicacy.

I.4 A Process-Oriented Model of Graphicacy

The Processes-Oriented Model of Graphicacy (POMoG) links graphicacy and graph comprehension research. On one hand, the POMoG acknowledges the task demands proposed by graphicacy frameworks, and the item difficulty and influencing factors from competence modeling. On the other hand, it also includes the comprehension processes underlying the construction of internal representations investigated in graph comprehension research. The POMoG has five core assumptions: (1) individual differences in graphicacy are manifested in differences in comprehension processing, (2) comprehension processes lead to the construction of internal representations, (3) an internal representation of the task, the graph, and the content is required to solve a task, (4) individual and task characteristics determine the process components of the comprehension process, (5) process measures are indirect indicators of comprehension processes. Figure 2 displays the POMoG, which consists of four components: comprehension process, internal representations, individual characteristics and task characteristics. The following sections define the four components of the POMoG (see Table 4 for an overview) and explain the model's architecture (Figure 2).

I.4.1 Model components

Individual characteristics are individual dispositions that are stable in the medium term at least. Graph schemata, domain knowledge, mathematical knowledge, reading comprehension, and general cognitive abilities, as well as age, grade, and gender are all individual characteristics. Individual characteristics influence comprehension processes and the construction of internal representations.

C*omprehension processes* describe what individuals do to construct internal representations. Comprehension processes include visual, visual imagery and mental processes, each of which consists of multiple process components. Visual processes are visual search, mapping, encoding, comparing and fluent reading. Visual processes are manifested in behavior, for instance in eye movements. Visual imagery process are mental manipulations of elements of the graph, for instance, using the mental image to compare the height of bars with different origins. Mental processes include other manipulations of information, for instance, performing mathematical operations, making inferences, or integrating information. Both visual imagery and mental processes are necessary whenever one's task goal requires information that is not directly displayed in the graph. In contrast to visual processes, visual imagery and mental processes are not directly manifested in a certain behavior (see Table 3).

Internal representations are mental states that allow individuals to answer graph comprehension tasks correctly. Graph comprehension requires three internal representations (IR). First, IR of the task i.e., understanding what the task requires; second, IR of the graph, i.e., understanding how the data is displayed; and finally IR of the graph's content, i.e., understanding what the data means. A graph comprehension task usually requires all three internal representations; if one is missing or incorrect, the graph comprehension task will not be solved correctly. Notably, the IR of the graph in the POMoG is analogous to the IR of the display in Shah et al. (2005). The term 'graph' is used instead of 'display' because the POMoG specifically deals with graphs. Moreover, the IR of the content in the POMoG is analogous to the IR of the referent in Shah et al. (2005). 'Content' is used instead of 'referent' because content is more concrete.

Item characteristics are features of an item that can be quantified via cognitive task analysis (Korossy, 1999). Item characteristics are referred to here in the general sense of stimulus characteristics. Item characteristics can be related to the task (e.g., whether one has to extract a single point or a relationship), the graph (e.g., line or bar chart) or content (e.g., simple vs. complex data structure, financial vs. biological data). Item characteristics can be systematically manipulated in an experiment or can be features of a real-world problem.

Table 3. Comprehension processes of the POMoG and their process components with definitions.

Comprehension processes	Definitions
Visual Processes	
Visual search	Search for indicators or labels[1]
Encoding	Read value of indicator or label[1]
Mapping	Attribute labels to axes, labels to indicators and values to indicators[1]
Comparison	Identify taller or longer indicator[1]
Spatial difference	Determine space in between indicators (length or height) [1]
Visual Imagery Processes	
Image Move	Move mental image of indicator on x or y axis[1]
Image Rotate	Move mental image of indicator around a point[1]
Image Compare	Compare mentally manipulated indicators[1]
Image Difference	Determine differences between mentally manipulated indicators[1]
Mental Processes	
Arithmetic operations	Addition, subtraction, division, multiplication of values[1]
Integration	Attribute pattern to an IR in a different representation[3]
Inference	Augment IR with information from prior knowledge[2]
Elaboration	Augment IR with new information[2]

[1]Gillan (2009), [2]Shah et al. (2005), [3]Lachmayer (2008)

I.4.2 Model architecture

The POMoG is 'process-oriented' because it explains item responses on the basis of the comprehension process between item exposure and response (see Figure 2). The POMoG assumes that three processing cycles take place between item exposure and item response. 'Cycle' here refers to construction-integration cycles in which meaning is constructed through word activation (e. g., words in sentences and labels in graphs), the formation of propositions, and the production of inferences and elaborations. A construction-integration cycle results in an interrelated network of units. This network can be integrated into prior knowledge (Kintsch, 1988). The POMoG distinguishes between three cycles because each cycle addresses different internal representations, involves different comprehension processes, and requires different prior knowledge. The construction of the IR of the task involves reading as the main comprehension process, as well as the individual characteristic of knowledge about word meanings. In contrast, the construction of the IR of the graph involves visual processes and knowledge of graphical conventions. Finally, the construction of the IR of the content can involve visual, visual imagery and other mental processes and content knowledge.

The role of the three IRs can be illustrated with an example item from the 2011 TIMSS assessment (see Figure 3). The figure shows 'population pyramids' for two countries denoted X and Y. The questions to be answered is: 'Why could the age structure of country X lead to more rapid population growth than the age structure of country Y?' In this example item, the IR of the task involves the concepts 'young people still get children', 'age structure is related to growth', and 'only young individuals contribute to population growth'. Individuals need to read the question and infer what information they need to read from the graph. Making such an inference requires the activation of mathematical knowledge (e.g., growth, relative and absolute magnitudes) and background knowledge (e.g., young people lead to more growth).

The IR of the task is novel from a graph comprehension perspective; however, it is a separate IR in the POMoG because graphicacy frameworks describe a number of task demands that do not fall within the purview of the IR of the graph or the IR of the content (e.g., Friel et al., 2001). For instance, individuals are only able to answer the question 'Does 'A' increase linearly over time?' if they are familiar with the term 'linear'. Even though a person may understand the graph and be able to extract the necessary information, they will not be able to answer a graph comprehension task when the term 'linear' is unknown. Therefore, terminology is an important factor for difficulty. Another factor of difficulty is the propositional complexity of the task. For instance, the question 'What is the value of car A' [WHAT IS(VALUE, CAR A] has a low number of propositions, the question 'What is the value of red cars that drove more than 10000 kilometers in 1990' [WHAT IS(VALUE, CAR), BEING(CAR, RED), MORE(CAR, 1000KM), IN(CAR, 1990)] has a high number of propositions. Additionally, individuals can be more or less fluent in reading. Whether the comprehension process leads to the construction of an IR of the task depends on the reading fluency of the individual as well as the terminology and prepositional complexity of the task.

The IR of the display includes the concepts 'two countries in two graphs', 'age is mapped to relative populations', 'males are left and female are right', and 'three zones show different age groups'. Individuals map the indicator age to the y-axis, map the percentage of population to the x-axis, map sex to the two sides of the graph, recognize the redundancy between the y-axis position and the color of the sections, and recognize the equivalent structure of both graphs. Individuals' knowledge of graphical conventions may help them understand the structure of the graphs.

Knowledge of graphical conventions and graph type can influence which comprehension processes are applied. For instance, individuals with knowledge of conventions of histograms do not need to encode the y-axis of the histogram, because histograms by definition have frequency on the y-axis. In this case, knowledge of graphical conventions reduces the number of comprehension process involved in the construction of an IR of the graph. In contrast, some graph characteristics generally increase the number of required comprehension processes. More mapping is required to construct an IR of the graph when the graph has low spatial compatibility (Huestegge & Philipp, 2011), the graph type is unfamiliar (Rose graph; Wainer, 1992), or the graph has high dimensionality (Shah et al., 2005). When the graph has irrelevant or misleading design features (Hullman, Adar, & Shah, 2011), visual searches become more demanding.

Finally, the IR of the content in the example item involves the notion ‘country X has a greater proportion of old people compared to young people’. Therefore, test takers need to compare the relative share of the population between ages 0 – 19 across graphs to answer the task. Generally, the IR of the content is constructed by searching for indicators and labels, encoding them and mapping them to their axis positions, and applying visual imagery and mental processes. The construction of the IR of the content becomes more demanding as the number of comprehension processes required increases. Moreover, mental processes tend to be more demanding than visual imagery processes, and visual imagery more demanding than visual processes (Gillan, 2009). For instance, comparing the height of two bars with different origins is challenging because rather than visually comparing them, one has to use imagery, i.e. proportional judgment (Hollands & Spence, 1998). Furthermore, a low cognitive fit between the task and graphs (Hegarty & Just, 1993; Shah et al., 1999) increases the number of mappings and comparisons required to construct an IR of the content. Similarly, more (visual) mapping is required when the graph has a weak spatial-to-conceptual mapping (Okan, Garcia-Retamero, Galesic, Cokely, & Maldonado, 2012). Predictions and interpolations requires individuals to make inferences based on their knowledge about functions and to use imagery to aggregate data points, respectively (Friel et al., 2001; Lachmayer, 2008).

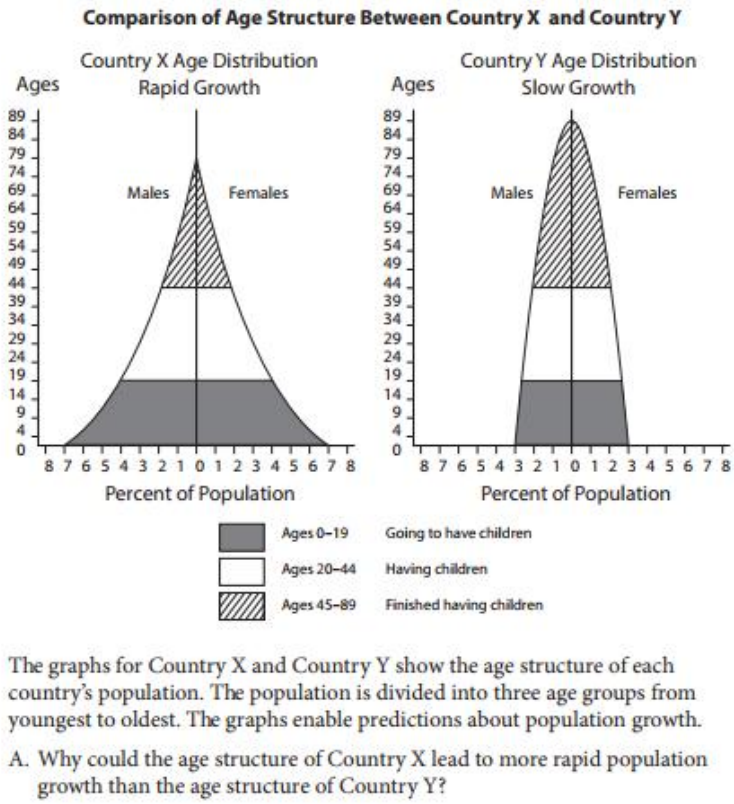


Figure 2. Example item from the Trends in International Mathematics and Science Study

Up until now, this section has mostly explained how task characteristics make IR construction more demanding. Individual characteristics that facilitate IR construction represent the opposite side of this. Specifically, there are three mechanisms underlying the influence of individual characteristics on IR construction. First, individual characteristics can substitute for process components; second, they can increase the fluency of process components; and third, they can control the usage of process components. For instance, knowledge of graphical conventions can substitute for mapping processes. When individuals are familiar with a graph type, they need to perform fewer mapping processes to make sense of the graph (Hegarty, 2013). An example of the second mechanism is arithmetic fluency accelerating arithmetic operations. Notably, in distinction to the substitution mechanism, individuals still perform the same process component; however, the arithmetic operations are less demanding for individuals with greater arithmetic fluency (Geary, Frensch, & Wiley, 1993). An example of individual characteristics controlling the usage of process components was given by Gillan (1995). In this study, individuals had to extract the mean from graphs with multiple bars. This task can be solved either by mentally calculating the arithmetic mean using the bar's values or by estimating the spatial mean between the bars without encoding the values of the individual bars. Individuals who were aware of the second strategy performed the task faster. Therefore, knowledge of this strategy made individuals use different process components. Interestingly, individuals applying this strategy were not just faster, they were also able to solve tasks with more bars and values. Similarly to this example, content knowledge can aid comprehension processes not just by substituting for, but also by controlling the comprehension process (Shah & Freedman, 2011; Cox, Romero, du Boulay, & Lutz, 2004).

To summarize, each of the three mechanisms can potentially aid IR construction. However, substitution and fluency aid IR construction by shortening the overall comprehension process, which implies that IR construction is performed faster. In contrast, choosing process components can either mean performing faster or performing IR construction at all. The differences between the three influencing mechanisms of individual characteristics should become evident in the relationship between comprehension success and process measures (e.g., time-on-task)

Goldhammer et al. (2014) showed that time-on-task and task success in an information literacy assessment are negatively related (see also Naumann & Goldhammer, 2017). The authors argued that individuals can be both faster and more accurate when the tasks requires automatic processing. Therefore, individuals who are fluent in the individual process components can be fast and accurate at the same time. In their case, fluency referred to process components of reading such as phonological recoding, orthographic comparison, and the retrieval of word meanings from long-term memory (Richter, Isberner, Naumann, & Neeb, 2013). However, they also showed that the relationship between time-on-task and task success becomes less negative as tasks become more complex. They argued that the less negative relationship indicates that comprehension processes need to be more controlled at high levels of complexity (Goldhammer et al., 2014). In other words, as tasks get more complex, the fluency of process components becomes less important relative to the control of process components.

The differences between the three mechanisms can be further illustrated using a physical analogy. Van Der Linden (2009) pointed out that the speed equation, i.e., ‘rate of change of some measure with respect to time (p. 258),’ should be applied to psychological processes. In this analogy, individuals solving test items are considered as analogous to individuals who have to travel from a starting point to an end point. Item complexity is the ‘straight-line distance’ between the starting (i.e., item exposure) and ending points (i.e., item solved). Individuals generally have to travel further to solve more complex items. Additionally, individuals may or may not reach the endpoint (i.e., solve or not solve the item). Individuals who engage in substitution characteristics can start closer to the endpoint. Individuals can be more or less fluent in process components, meaning that they travel faster. Individuals can control the process components more or less effectively, meaning that the ‘distance traveled’ can deviate from the straight-line distance depending on how directly individuals navigate. In turn, the speed with which individuals reach the endpoint is a combination of where they start, how fast they travel and how directly they navigate. Comprehension processes are determined by individual characteristics related to substitution, fluency and control. Notably, navigation becomes more important as the distance increases, just like complex tasks require more controlled processing (Goldhammer et al., 2014). Fluency is most important for simple tasks in which the endpoint is in ‘viewing distance’. Furthermore, this analogy illustrates the two main aspects that need to be considered when inferences about comprehension processes have to be made.

First, inferences about comprehension process have to be made based on the available information. In empirical studies, only item response and process measures can be observed. In the traveling analogy, individuals’ overall speed can only be compared based on their times when the distance is the same for everyone. Speed is distance divided by time; when the distances are the same, time equals speed. Analogously, process measures can be linked to comprehension processes if everyone solves the same items. For instance, time-on-task is an indicator of processing speed if everyone solves the same items. However, graphicacy tests are intentionally constructed so that not everyone solves every item. Therefore, process measures are not directly linked to comprehension processes in graphicacy tests. Just like time and distance, comprehension processes reflect the relationship between process measures and comprehension success. Building on this logic, the influence of individual characteristics for comprehension processes can be investigated based on how they change the relationship between process measures and comprehension success.

Second, the analogy states that navigation and speed are crucial. A mindful traveler will first explore their surroundings in order to navigate more effectively, and then approach the goal at full speed. A similar approach has been described in comprehension research. Schnotz et al., (2014) distinguished between two phases of text-picture comprehension: initial reading and task completion (see also Schnotz & Wagner, 2018). During initial reading, processing is coherence-oriented and general. The emphasis is on a global understanding of the content, which includes the construction of an internal representation that does not concern the specific task. Viewers process the material in order to understand the subject matter as a whole. During task completion, processing is goal-oriented and selective for task-relevant information. Viewers focus primarily on the task to be solved and select the relevant information from the external source of information.

Similar distinctions were made by Lindner Eitel, Strobel, and Köller (2017), who found that students mainly focus on the question at the beginning and the answer options towards the end of the decision-making process. They divided item response into two phases: (1) an information acquisition phase, in which students construct a mental representation of the problem or situation described in the item stem, and (2) a decision-making phase, in which students evaluate the answer options with respect to the problem to be solved (Lindner et al., 2017; Greiff et al., 2013).

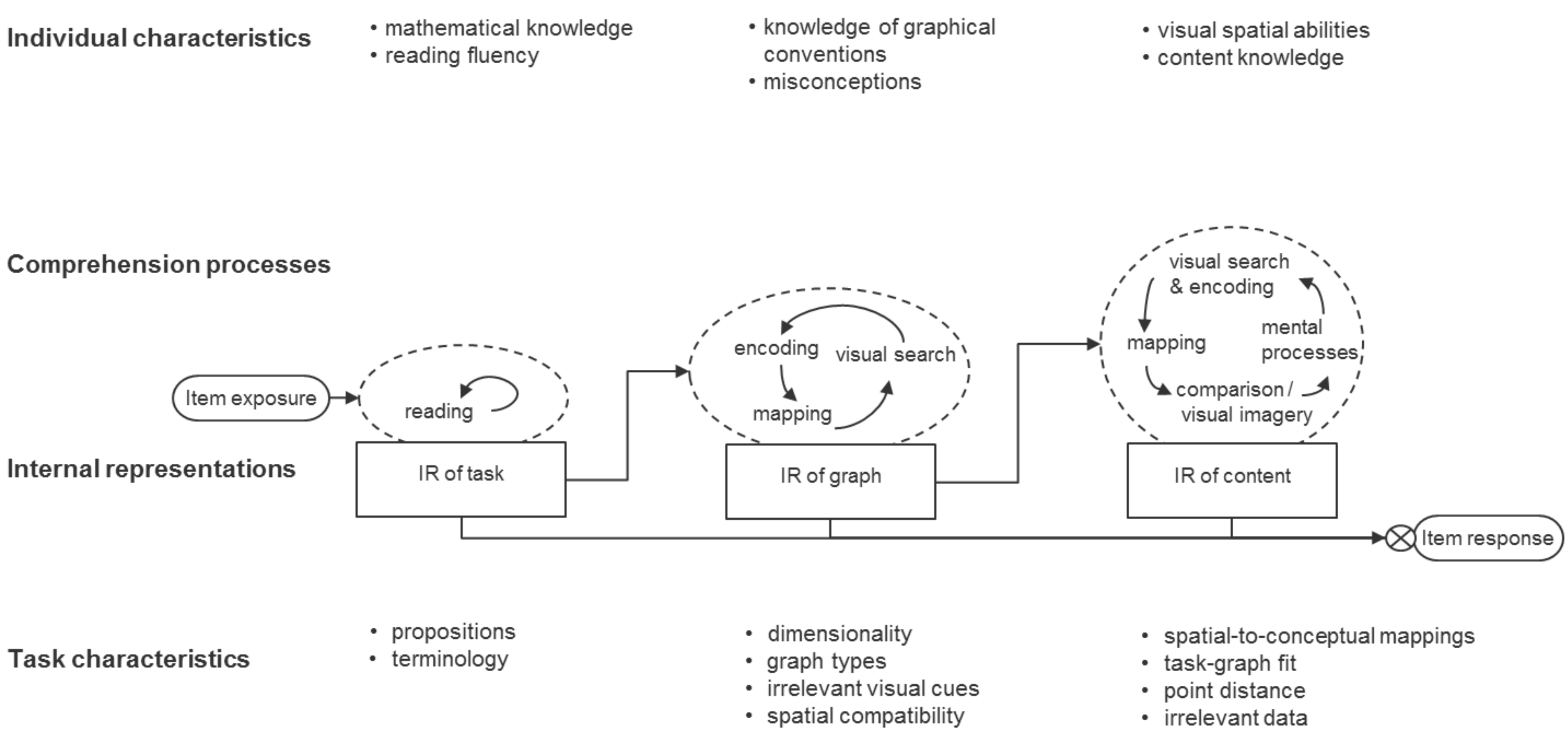


Figure 3. Diagram summarizing the Process-Oriented Model of Graphicacy (POMoG)

I.5 Research questions

Up until now, this thesis has developed the POMoG to combine research on graphs from the literacy and comprehension communities. The proposed model combines the 'real-world' task demands of graphicacy research with models on comprehension processes from graph comprehension research. The POMoG explains item responses in graphicacy tests on the basis of individual and item characteristics, comprehension processes, and the presence of internal representations. Additionally, it proposes a rationale for interpreting process measures as comprehension processes.

The POMoG has five core assumptions: first, individual differences in graphicacy are manifested in differences in comprehension processes. Second, comprehension processes lead to the construction of three internal representations. Third, internal representations of the task, the graph, and the content are required to solve a task. Four, individual and task characteristics determine the process components of the comprehension process. Five, process measures are indirect indicators of comprehension processes. Additionally, individual characteristics can influence graphicacy via three different mechanisms, by substituting for process components, by making process components more fluent, and by controlling the usage of process components. Based on a travel analogy, the POMoG assumes that the relationship between process measures and comprehension success serves to represent comprehension processes.

The following studies address three of the core assumptions of the POMoG. The first study (Chapter II) investigates the influence of basic numerical abilities on graph reading performance. The influence of basic numerical abilities on graph reading performance can explain the underlying comprehension process because specific basic numerical abilities can be linked to specific process components of graphicacy performance. The study distinguishes among eight basic numerical abilities: addition, subtraction, multiplication, number line estimation, proximal arithmetic skills, conceptual knowledge about arithmetic operations, basic geometry, and non-symbolic magnitude comparisons. These basic numerical abilities can be linked to processing components of the POMoG. For instance, number line estimation can be linked to comparison processes. Therefore, the influence of basic numerical abilities addresses the core assumption that individual differences in graphicacy are manifested in comprehension processes. Consequently, the first research question is 'how do basic numerical abilities influence graph reading?'

The second study (Chapter III) addresses the fifth core assumption, 'process measures are indirect indicators of comprehension processes'. The POMoG argues that the relationship between process measures and comprehension success is representative of comprehension processes. However, it also states that comprehension processes change as a function of the comprehension phase (initial reading vs. task completion phase). Subsequently, it is hypothesized that the relationship between process measures and comprehension success depends on the comprehension phase. The study focus on the process measures ‘text-graph transitions’ and ‘time-on-task’ as indicators of ‘integration’ because these measures are well established in multimedia research. Interestingly, text-graph transitions allow for two contradicting interpretations. Text-graph transitions can either indicate the integration of information from the graph and text or disorientation, the inability to find relevant information. Furthermore, individuals’ reading speed, graph comprehension ability and content knowledge are assumed to facilitate the comprehension of graph and text. The POMoG assumes that these individual characteristics substitute for process components, make them more fluent, or help control the comprehension process. In short, the POMoG stats that individual characteristics influence the comprehension process and therefore influence the relationship between text-graph transitions, time-on-task and comprehension success. Consequently, the Chapter III addresses two research questions: ‘Does the relationship between process measures and comprehension success depend on the comprehension phase?’ and ‘Do individual characteristics influence the relationship between process measures and comprehension success?’

The third study (Chapter IV) addresses the POMoG’s assumption that ‘internal representations of the task, the graph, and the content are required to solve a task’, and specifically the hierarchical dependency between the IR of the graph and content. Such a hierarchical dependency between these internal representations is linked to a debate in multimedia research. Specifically, two contradicting perspectives within multimedia research describe the hierarchical dependency between the comprehension processes that need to be performed for integrated text-graph comprehension. The text-centered perspective states that the IR of the content is a prerequisite for the IR of the graph. Conversely, the multiple-representation perspective implies that the IR of the graph is a prerequisite for the IR of the content. Therefore, a multimedia paradigm is used to investigate the hierarchical dependency between these two internal representations, comparing the two contradicting perspectives. This leads to the research question: ‘Is a text-centered or multiple-representation perspective more applicable to text-graph comprehension?’

Notably, the questions in Chapters III and IV cannot be answered with standard statistical modeling approaches. Therefore, the following section introduces common modeling approaches and evaluates their ability to address the research questions posed in Chapters III and IV.

I.6 Statistical modeling approaches

Graphicacy has been investigated using competence modeling approaches such as item response theory (e.g., Lachmayer, 2008; Lai et al., 2016). However, this thesis proposes a process-oriented model of graphicacy. Therefore, this section discusses common modeling approaches and evaluates their ability to capture the comprehension processes of graphicacy. Numerous aspects need to be considered when selecting an appropriate modeling approach. A modeling approach can conceptualize competence as a continuous dimension or as discrete states, and they should be able to account for the influence of process measures as well as individual and item characteristics. Each modeling approach has psychometric implications, practical implications, and enables different inferences about underlying comprehension processes.

The most prevalent modeling approaches in educational research are based on item response theory (IRT). IRT conceptualizes competence as a numerical continuum (Walker & Beretvas, 2003). The most commonly used model in IRT is the Rasch model (Wu & Adams, 2007). The basic assumption of the Rasch model is that 'the probability of a correct item response only depends on one's individual ability and item difficulties'. The Rasch model is used in many large-scale assessments because it has several psychometric and practical advantages, for instance, interval-scaled measurement, a joint scale for items and individuals, and easy reliability estimation (Wright, 1977). The Rasch model is designed to estimate a numerical value that accurately represents individual competence. This numerical value can be used to compare individuals; however, it carries no information about what distinguishes individuals and what level of performance individuals are capable of. Therefore, Rasch models do not allow for inferences about learning or comprehension processes (Leuders & Sodian, 2013; Rupp & Mislevy, 2007). Furthermore, the assumption that 'the probability of an correct item response depends only on one's individual ability and item difficulties' is very restrictive because in many domains it is plausible that individuals' performance depends on multiple competencies. Additionally, in the Rasch model item difficulty is random and not explained by item characteristics. However, these aspects have been addressed in different modeling approaches, namely multi-dimensional item response theory (MIRT) and the linear logistic test model (LLTM).

MIRT accounts for the possibility that item responses are the result of multiple competence dimensions (Hartig & Höhler, 2008). The assumption is that 'the probability of a correct item response depends on (multiple) individual abilities and item difficulties'. This modeling approach can account for relationships between different competence dimensions, and even can account for items that require a mixture of abilities (Klieme, Hartig, & Rauch, 2008). This modeling approach enables inferences about comprehension processes on the basis of which items required similar or different competencies. Furthermore, when individuals' performance is the result of multiple competence dimensions, it is sometimes inferred that different competencies are associated with different processing systems (Schroeders, Bucholtz, Formazin, & Wilhelm, 2013). However, MIRT still does not allow for inferences about what performance individuals are actually capable of.

What performance individuals are capable of is addressed by Fischer's LLTM (1995), which explains item difficulty based on item characteristics. The model is similar to the Rasch model except that item difficulty is explained by item characteristics. The assumption is that 'the probability of a correct item response depends on one's individual ability and the items' characteristics'. This modeling approach allows for inferences about comprehension processes because an individuals' competence should be related to the task demands she or he is capable of performing (Embretson & Daniel, 2008).

Another way of enabling inferences about comprehension processes is to include process measures. Goldhammer et al. (2014) presented a time-on-task model in which they assumed that 'the probability of a correct item response depends on time-on-task, individual ability, the individual's processing speed and item difficulties'. This modeling approach allows inferences to be made about the comprehension process based on the strength of the correlation between time-on-task and task success. For instance, they show that the correlation is positive in problem-solving tasks and negative in literacy tasks. However, in this modeling approach no individual or item characteristics explain the strength of the correlation.

The Rasch, MIRT, LLTM, and time-on-task models all conceptualize competence as a numerical continuum. Yet in many cases learning and competence development is non-continuous, especially when 'conceptual change' is involved (Leuders & Sodian, 2013). When competence development is non-continuous, individuals should be evaluated with regard to their competence 'state' or 'stage' rather than their position on a continuum. Cognitive diagnostic models (CDM; De La Torre, 2011) and knowledge space theory (KST; Falmagne, Koppen, Villano, Doignon, & Johannesen, 1990) address discontinuous competence development. CDMs assume that 'the probability of a correct item response depends on a discrete set of skills (and slipping and guessing error)'. The CDM allows inferences to be made about comprehension processes because the assumed skill set is directly related to item characteristics (similar to the LLTM). An individual's skill set refers to the task demands the individual can master. KST functions similarly, except that it additionally accounts for hierarchical dependency between items. KST was extended to estimate probabilistic knowledge structures with basic local independence models (BLIM). A BLIM assumes that 'the probability of a correct item response depends on individuals' knowledge state within a knowledge structure (and slipping and guessing errors)'. KST and BLIMs allow inferences to be made about the comprehension process based on the hierarchical dependencies between knowledge states represented by the knowledge structure. The knowledge state in a knowledge structure refers to what performance an individual is capable of, what other abilities this implies, and what the individuals is able to learn in the future.

To sum up, common modeling approaches were reviewed and evaluated with regard to their ability to allow inferences to be made about the comprehension process. First of all, whether competence is seen as a continuous dimension or a discrete set of skills is essential. Generally, it is more informative to use discrete modeling approaches (i.e., CDM and KST) when competence development involves 'conceptual change'. Otherwise, it is more practical to conceptualize competence as a continuum (i.e., Rasch, MIRT, and LLTM), especially when items involve a variety of different comprehension processes. Second, a modeling approach allows inferences to be made about comprehension processes when it accounts for item characteristics (i.e., LLTM, CDM, and KST). When a model accounts for the influence of item characteristics on item difficulty, individuals' competencies are associated with the task demands they are capable of performing. Additionally, modeling approaches can account for the influence of process measures on tasks to enable inferences about what comprehension processes lead to task success.

In light of all this, what is a suitable modeling approach to investigate the research questions posed here? In Chapter III, which concerns the questions 'Does the relationship between process measures and comprehension success depend on the comprehension phase?' and 'Do individual characteristics influence the relationship between process measures and comprehension success?', text-graph comprehension is understood as a continuous competence. However, the chapter additionally addresses the influence of time-on-task and text-graph transitions on comprehension success and how this influence changes across comprehension phases and in association with individual characteristics. These aspects are not addressed by any single one of the presented modeling approaches; however, the modeling approaches can be combined. Therefore, an IRT-based model with explanatory variables and process measures variables was developed to address the research questions posed in Chapter III. This model, the Explanatory Model of Person-Process Interaction (EMPPI), is presented in the following section.

Next, Chapter IV addresses the question 'Is a text-centered or multiple representation perspective more applicable to text-graph comprehension?' The study compares two perspectives that assume contradicting hierarchical dependencies between comprehension processes during the comprehension of text and graphs. KST is the only modeling approach that can address hierarchical dependencies between comprehension processes. Therefore, KST and more specifically the BLIM will be applied in Chapter IV.

The next section addresses the two modeling approaches, EMPPI and BLIM. The EMPPI was developed specifically to investigate the questions from Chapter III. The BLIM is an already existing modeling approach. However, the BLIM is nevertheless introduced because it is currently mostly discussed in methodological journals.

I.6.1 Item response theory with explanatory and process variables

This section develops an IRT that accounts for the influence of explanatory variables and process variables. In the previous section, it was concluded that modeling approaches have to account for item characteristics and/or process measures in order to be able to make inferences about comprehension processes. Wilson and De Boeck (2004) distinguished between descriptive and explanatory item response models. Descriptive item response models focus on measurement issues, which can help to improve the quality of tests in terms of reliability, but do not contribute to construct validity. Reliability concerns how precisely a test measures a given competence, while construct validity concerns what the test actually measures. Wilson and De Boeck (2004) suggested including 'explanatory variables' to improve construct validity. Explanatory variables can be either individual or item characteristics. Individual characteristics explain why some individuals are more competent than others, while item characteristics explain why some items are more difficult than others. For instance, an (explanatory) individual characteristic for graphicacy would be reading speed, and an explanatory item characteristic would be the propositional complexity of the questions. In the sense they are used here, individual and item characteristics are always explanatory. Therefore, they are simply referred to as individual and item characteristics. One modeling approach that accounts for both individual and item characteristics is the double explanatory item response model (Wilson & De Boeck, 2004).

The double explanatory item response model combines the LLTM, which explains item difficulty based on item characteristics, and a latent regression Rasch model (LRRM) that explains individuals' competencies based on individual characteristics. It is also referred to as a latent regression linear logistic test model (latent regression LLTM)[3]:

$$\eta_{pi} = \sum_{j=1}^{J} \vartheta_j Z_{pj} + \varepsilon_p - \sum_{k=0}^{K} \beta_p X_{ik} \qquad \text{F1}$$

The latent regression LLTM explains item responses as the result of a combination of item and individual characteristics. The item characteristics have a fixed effect β_j and individual characteristics have a fixed effect ϑ_j. The item characteristics also include a constant predictor β_0, while the individuals' effect is composed of the fixed effect and the residual term ε_p. For graphicacy, typical item characteristics could be level of complexity, graph type or graph-task fit. Individual characteristics might be reading comprehension, knowledge of graph conventions or spatial abilities. For instance, $\beta_{complexity}$ represents the effect of the level of complexity and $\vartheta_{reading}$ the effect of reading comprehension on the probability of correct responses, whereas ε_p represents individual variation in item responses that cannot be explained by reading comprehension.

The latent regression LLTM accounts for individual and item characteristics. However, to address the research questions from Chapter III, the model additionally needs to account for the influence of process measures. Therefore, process measures are included as item-by-person variables. Essentially, the time-on-task model by Goldhammer et al. (2014) is integrated with the latent regression LLTM (Wilson & De Boeck, 2004). In addition, Chapter III addresses the influence of individual characteristics on the relationship between process measures and task success. Therefore, the model has to account for the interaction between individual characteristics and process measures. The resulting model is the Explanatory Model of Process-Person Interaction (EMPPI). Below, the EMPPI is constructed in a step-by-step process beginning with the latent regression LLTM. .

First, process measures (W_{piH}) are included as a fixed effect on the item-by-person level. The fixed effect of a process measure (δ_H) depends on general nature of the task. For instance, Goldhammer et al. 2014 show that the fixed effect of time-on-task is positive for problem-solving

[3] *Formula 1: The latent regression linear logistic test model defines the probability of a correct answer π_{pi}~η_{pi} as the sum of individual characteristics effects ϑ_j, a residual term ε_p, and the sum of item characteristics effects β_k. Z_{pj} is the matrix of individuals and their characteristics and X_{ik} is the matrix of items and their characteristics.*

tasks and negative for literacy tasks. However, to investigate individual differences in comprehension processes, process measures need to be included as individual random slopes (ζ_{ph}). These random slopes represent how much the individual effect of the process measure deviates from the fixed effect of the process measure. For instance, when the individual effect is more negative than the fixed time-on-task effect, it means that individual solves items faster than other individuals. In this case, the individual effect represents the individual's processing speed (Van Der Linden, 2009).

In the EMPPI, the individual random slopes for the process measures can be explained by the fixed interaction effect (ψ_{JH}) of a process measure (W_{piH}) and individual characteristics (Z_{pJ}). The fixed interaction effect represents the extent to which an individual characteristic changes the individual effect. For instance, reading speed may decrease the time it takes to solve a task. In other words, reading speed increases processing speed in graphicacy. When the process-person interaction is included, ζ_p represents a residual term, or the variation in the individual effect of process measures that cannot be explained by the process-person interaction. The formula F2 summarizes the different components and Figure 3 presents a graphical representation of the EMPPI.

$$\eta_{pi} = \sum_{j=1}^{J} \vartheta_j Z_{pi} + \varepsilon_p + \sum_{h=1}^{H} \delta_h W_{piH} + \sum_{j \in J, h \in H}^{JH} \psi_{hj} W_{pih} Z_{pij} + \xi_p - \sum_{k=0}^{K} \beta_k X_{ik} \qquad \text{F2}$$

The EMPPI combines different modeling approaches to maximize the ability to make inferences about underlying comprehension processes that constitute individuals' performance. In the case of graphicacy, first, individual characteristics can explain which other abilities are associated with graphicacy. Second, item characteristics explain why some items are more difficult than others. Third, process measures explain what behaviors are associated with comprehension success. Fourth, the person-process interaction explains the difference in the individual effect of process measures. It therefore covers all aspects that will be investigated in Chapter III.

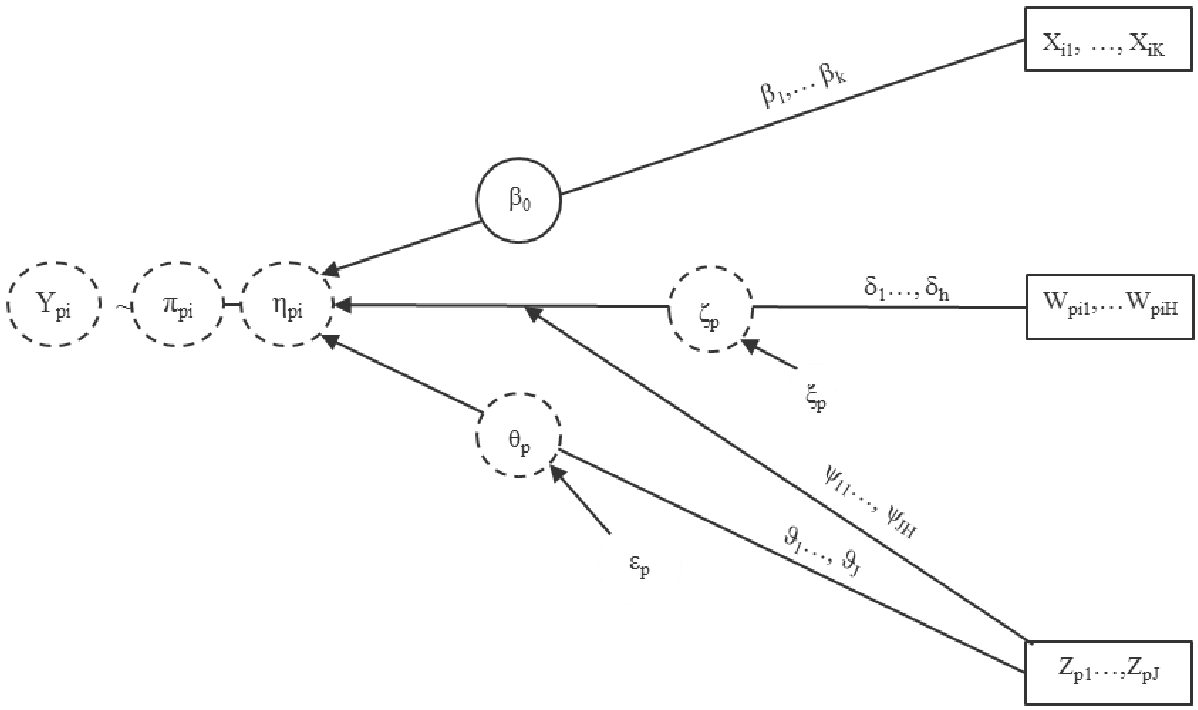


Figure 4. Graphical representation of the Explanatory Model of Person-Process Interaction (EMPPI)

Data:

Y_{pi} denotes the response of a person p to item i;

Indices:

p for person

i for items

K item predictor

J for person predictors

H for process measures (person-by-item) predictors

Predictors:

X for item predictors (X_{iK})

Z for person predictors (Z_{pJ})

W for person-by-item predictor (W_{piH})

Model:

πpi for the probability of $Ypi = 1$

η_{pi} for the transformed πpi based on the link function

Effects:

θ for person random intercept (θ_p)

ζ_{pH} for person random slope of person-by-item predictors (ζ_p)

β for fixed effects of item predictors (β_K)

ϑ for fixed effects of person predictors (ϑ_J)

δ for fixed effects of person-by-item predictors (δ_H)

ψ for fixed effect of process-person interaction (ψ_{JH})

ε person residual when predictors included

I.6.2 Knowledge Space Theory

Knowledge space theory (KST; Falmagne et al., 1990) was developed to assess students' knowledge of mathematical and science concepts. According to KST, all knowledge in a certain domain has a domain ontology (Heller, Steiner, Hockemeyer, & Albert, 2006). The domain ontology describes the concepts within the domain and how they depend on each other. For instance, the domain 'fractions' may consist of multiplying, dividing, and adding fractions and finding a common denominator. The concepts related to fractions depend hierarchically on each other. For instance, adding fractions requires students to find a common denominator. KST is a competence modeling approach that takes this hierarchical dependency into account. The classical example from Falmagne et al. (1990) illustrates how KST determines the dependency between concepts using five mathematical problems (left panel, figure 4).

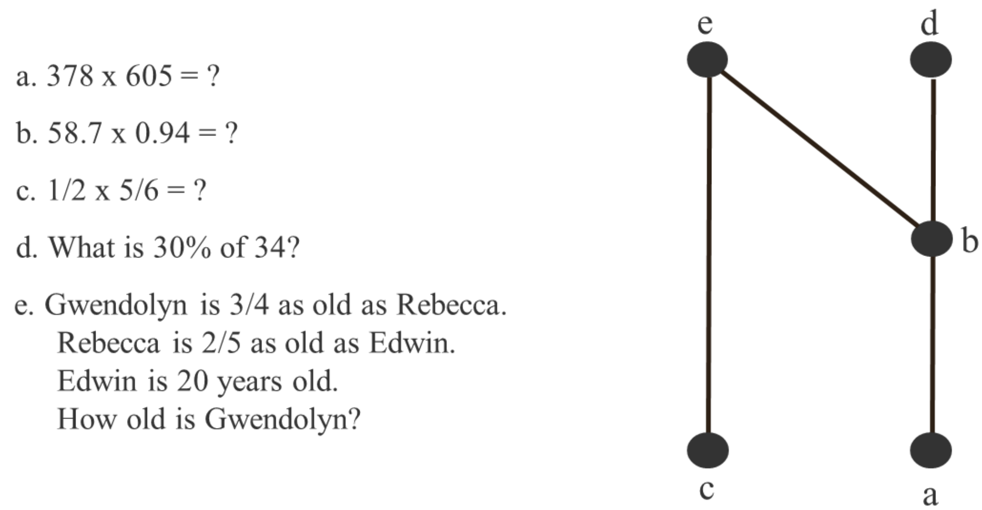


Figure 5. Five problem from Falmagne et al 1990 (left panel) and Hasse diagram of the knowledge structure (right panel).

The five mathematical problems represent the knowledge domain. The problems involve different combinations of mathematical concepts. For instance, to answer Problem d (i.e., what is 30% of 34?) student need to multiply an integer with a decimal number (i.e., 34 x 0.3). Students who manage to solve Problem d should be able to multiply integers and multiply decimal numbers. Therefore, students who get Problem d correct should get the first (i.e., a. 378 x 606) and second (b. 58.7 x 0.94) problems correct as well. This is an example of hierarchically dependency between mathematical concepts. Hierarchical dependencies are important for education because they determine which concept should be learned next. It is difficult to learn how to calculate percentages if one is not able to multiply decimals. In some cases, concepts even depend on multiple other concepts. In the example (Figure 4), Problem e requires the concepts of multiplying decimal numbers and multiplying fractions. All of these hierarchical dependencies[4] can be combined to form the knowledge structure of the domain. Figure 4 (right panel) shows a Hasse diagram that represents the knowledge structure of Falmagne et al. 's (1990) five example problems. The lines between the elements each represent a hierarchical dependency. In the following, it is explained how KST formulates knowledge structures.

Generally speaking, KST is a set-theoretical framework. Therefore, in contrast to other modeling approaches, KST is based on combinatorics. A knowledge structure delimits a space of all different possible combinations of responses. KST uses a notation system to define knowledge structures. A knowledge structure is defined as a pair (Q, K) in which Q is a non-empty set, and K is a family of subsets of Q. The set Q is called the domain of the knowledge structure. Q consists of elements that are referred to as *items*. The subsets of items in the family K are labeled knowledge states (Doignon & Falmagne, 1985). Items are denoted with parentheses (i.e., (a)), and knowledge states are denoted with brackets (i.e., {a}). A knowledge state represents the subset of items in the domain Q an individual has mastered.

The domain Q in the example in Figure 4 has five items. Each of the five items can either be solved or not solved; therefore, students can generate any of $2^5 = 32$ different response patterns. Hence, the knowledge structure of Q could contain up to 32 different knowledge states. However, not all response patterns are plausible due to the hierarchal dependencies between concepts. Following the knowledge structure in Figure 4, only ten different knowledge states are possible:

$$Q = \{\{\varnothing\},\{a\},\{c\},\{b,c\},\{a,c\},\{a,b\},\{a,b,c\},\{a,b,d\},\{a,b,c,e\},\{Q\}\} \qquad \text{F3}$$

Only ten knowledge states are plausible because (d) depends on (c) and (b), and (b) depends on (a). Thus, knowledge states that include (d), but not (c), (b), and (a) can be excluded. Also, all knowledge states that include (c), but not (b) and (a) can be excluded because (c) depends on (a) and (b).

---

[4] *Many equivalent terms are used to describe hierarchical dependencies: sub-sequential relationships, partial orders, prerequisites, sumise relationships, and hierarchical relationships.*

Until now, it was assumed that a student's response pattern is a direct reflection of their knowledge state. However, in practice, students may guess the correct answer on an item or fail to correctly answer an item due to a careless mistake. The basic local independence model (BLIM) can account for the influence of these response errors. The BLIM assigns observed response patterns to a probabilistic knowledge structure. Practically, the BLIM assigns responses that contradict the hierarchical dependency of the domain to a plausible knowledge state. For instance, a student who answers only problems (b) and (d) correctly can be assigned to the knowledge state {a,b,d} because it is likely that the student made a careless error in (a). In a BLIM, the rate of careless errors and guessing errors are assumed to be specific to individual items and the same for everyone.

Formally, a BLIM calculates the probability of the observed response pattern given the knowledge state of a knowledge structure, taking guessing and slipping error into consideration (Heller & Wickelmaier, 2013). The response error is defined for each item, such that the responses are stochastically independent over items q and the response to each item q only depends on the probabilities of β_q slipping, η_q guessing and a person's knowledge state K. The probability of the response pattern R given the knowledge state K is determined as follows (Heller & Wickelmaier, 2013):

$$P(R|K) = \prod_{q \epsilon K/R} \beta_q \prod_{q \epsilon K \cap R} 1 - \beta_q \prod_{q \epsilon K/R} \eta_q \prod_{q \epsilon Q\ (R \cup K)} 1 - \eta_q \qquad \text{F4}$$

The BLIM faces a specific estimation problem. The likelihood of any probabilistic knowledge structure increases with greater response error. Additionally, knowledge structures with few hierarchical dependencies are more likely than very hierarchical knowledge structures. However, hierarchical knowledge structures with few restrictions are less informative because they only describe general combinatorics possibilities (Heller & Wickelmaier, 2013). Therefore, the BLIM uses a minimum discrepancy maximum likelihood (MDML) estimation method to find the probabilistic knowledge structure that is most informative but has plausible response error. MDML uses a trade-off between the minimum discrepancy method, which optimizes model parameters by minimizing the number of expected response errors, and the maximum likelihood method, which optimizes model parameters by maximizing the likelihood of the model (Heller & Wickelmaier, 2013).

In sum, KST can be used to determine the knowledge structure of a domain. This knowledge structure is based on the domain ontology and delimits the combinatorial space of students' responses. Students' responses can include response error. The BLIM takes response error into account and estimates a probability knowledge structure. Up until now, it was only referred to concepts in the mathematical domain. In mathematics, the domain ontology is relatively clear and the dependency between concepts within the domain is evident. However, in Chapter IV, the question 'Is a text-centered or multiple representation perspective more applicable to text-graph comprehension?' will be addressed because the domain ontology of text-graph comprehension has been described from two different theoretical perspectives. The multiple-representation and the text-centered perspectives imply different hierarchical dependencies in text-graph comprehension. Therefore, in Chapter IV a BLIM is used to determine which knowledge structure is more applicable to text-graph comprehension.

I.7 Study overview

This thesis has presented a summary of the research on graphicacy and graph comprehension. Second, it has integrated these two research communities in the Process-Oriented Model of Graphicacy (POMoG). Third, research questions that address the core assumptions of the POMoG have been developed. Finally, appropriate modeling approaches have been selected.

The following studies aim to find evidence for the core assumptions of the POMoG by investigating the influence of basic numerical abilities and the effect of process measures during different comprehension phases, as well as by evaluating the applicability of two contradicting perspectives from multimedia research.

Chapter II seeks to find evidence for the assumption that comprehension processes express individual differences in graphicacy. The influence of basic numerical abilities on graph reading performance can elucidate the underlying comprehension process because basic numerical abilities are rooted in neuro and cognitive science. Therefore, specific basic numerical abilities can be linked to specific process components. To this end, the influence of basic numerical abilities and graph reading performance is determined via a multiple regression analysis and their relative contributions are determined with a relative weight analysis. Additionally, the influence of general cognitive ability, age, and gender are considered as control variables. The analyzed sample consisted of 750 German students (grades nine to eleven).

Chapter III seeks to find evidence for the assumption that process measures are indirect indicators of comprehension processes. The study investigates how inferences about the comprehension process can be made on the basis of process measures. In multimedia research, transitions between a text and a graphic can be interpreted in two opposing ways. Text-graphic transitions can be interpreted as the integration of information from the text and graph or as disorientation, the inability to find relevant information. The POMoG argues that comprehension processes are represented by the relationship between process measures and comprehension success. Additionally, it is hypothesized that comprehension processes are influenced by the comprehension phase. Comprehension is either coherence-oriented or task-selective. The relationship between process measures and comprehension outcomes should be positive during the coherence-oriented phase and negative during the task-selective phase because processing should be more controlled during the coherence-oriented phase. Furthermore, the influence of individuals' characteristics on the relationship between process measures and comprehension success is examined. Chapter III includes two studies which analyze time-on-task and text-graph transitions in a total of 77 university students who worked on twelve text-graph integration items. The analysis is conducted with the EMPPI, which was specifically developed for the study.

Chapter IV addresses the hierarchical dependency between the internal representation of a graph and the graph's content. A multimedia paradigm is used to investigate this dependency because two contradicting perspectives have been described that hypothesize different hierarchical dependencies between these comprehension processes. The text-centered perspective states that the internal representation of the content of the text is a prerequisite for the internal representation of the graph. Conversely, the multiple-representation perspective implies that the internal representation of the graph is a prerequisite for the internal representation of the content. In this study, the response patterns of 50 adults who answered a large number of text-graph integration items were analyzed. The fit between the response pattern and the probabilistic knowledge structures for the two perspectives are compared with a BLIM.

Chapter II. Influences of basic numerical abilities on graph reading performance

Understanding graphically presented information is an important aspect of modern mathematical and science literacy. In our study we investigated the influence of basic numeric abilities on students' ability answer mathematical tasks with graphically presented information. We analyzed data of 750 German students (grades 9 – 11) and evaluated the determinants of graph reading performance with multiple regression analysis using predictors of basic numeric abilities (such as number line estimation, basic arithmetic operations, etc.), considering also the influences of general cognitive ability, age, and gender. We found that number line estimation, subtraction, and conceptual knowledge were significant predictors of graph reading performance beyond the influences of general cognitive ability. This indicates that basic numeric abilities are still relevant for real-life problem solving in secondary school. We discuss possible mechanisms which directly (through respective arithmetic procedures) as well as indirectly (through mathematization of the problem) effectuate that basic numeric abilities graph reading performance.

II.1 Introduction

"Graphs, charts, cartograms, thematic maps, etc., are common tools for handling and communicating quantitative information in contemporary society." (Åberg-Bengtsson & Ottosson, 2006, p. 112). During the last ten years, quantitative data have played an increasingly important role in our life. It has become a daily routine to interpret the line diagrams displaying our heart rate data when doing sports, to read bar graphs in newspapers and infographics on TV. It is thus uncontroversial that reading quantitative graphs is an important 21st-century skill (Ananiadou & Claro, 2009). Accordingly, many large-scale assessments of scholastic achievement refer to tasks involving quantitative graphs, for instance, within scientific literacy and mathematics tests in TIMSS (Baumert, Bos, & Watermann, 1998) or reading, mathematical and science literacy assessments in PISA (Deutsches & Baumert, 2013). Strikingly, 60% of the international 8 graders were able to read a single value of a line graph and only 29% were able to determine the average of a graph (TIMSS 2011 Assessment, 2013). Due to the omnipresence and relevance of graphs in everyday life and education, it is worth investigating the factors that influence students' ability to fluently extract and use information from graphs. This ability is closely related to graph literacy (Galesic & Garcia-Retamero, 2011).

The majority of research on graph reading addressed the influence of task (e.g., Lachmayer, 2008; Shah & Freedman, 2011) and graph (e.g., Wainer, 1992) characteristics (for reviews Friel, Curcio, & Bright, 2001; Shah & Freedman, 2011) on performance. In terms of the influence of graph reader characteristics, mathematical knowledge, experience with working on graphs, and general cognitive abilities have been identified as relevant factors for graph reading. However, to the best of our knowledge there is no research evaluating graph reading abilities from the numerical cognition perspective.

Importantly, however, investigating influences of basic numerical abilities on graph reading allows for interconnecting two fields of research: i) Graph reading research that focuses on the ability to perform authentic everyday tasks with graphs, and numerical cognition research which focuses on underlying processes, structure and development of basic numerical abilities (e.g., understanding number magnitude, basic arithmetical operations, etc.). Interestingly, graph reading tasks involve working with visually, spatially and symbolically-coded quantities. As such measures of basic numerical abilities may allow to further differentiate individuals' abilities to work with these different codes of quantity. Therefore, by investigating the influence of basic numerical abilities of graph reading, we may learn more about the underlying cognitive mechanisms that contribute to graph reading.

The current study addresses the question of how basic numerical abilities influence graph reading performance. In the following, we will give a brief review of the literature investigating students' graph literacy.

II.1.1 Graph reading

We define graph reading as the ability to fluently extract and use information from graphs. Graph reading is an important, if not the most essential aspect of graph literacy. Graph literacy is defined as the ability to understand information presented as graphs in everyday life (Galesic, & Garcia-Retamero, 2011). Various other terms have been used to denote similar constructs, such as graphicacy (Lowrie, Diezmann, & Logan, 2011; Åberg-Bengtsson, 2006), graphing ability (Berg & Smith, 1994) and graph sense (Friel et al., 2001).

Generally, graph reading involves decoding and interpretation of visually displayed information. For Lowrie and Diezmann (2011) decoding and interpretation of bar and line graphs requires knowledge about the graphical language. For bar and line graphs it is the apposed-position language which encodes information by a mark positioned along both the x and y-axes. To correctly interpret the coded information, it is necessary to integrate information from both axes. More specifically, Gillan and Lewis (1994) suggested that understanding a graph involves separate processing steps with nonarithmetic and arithmetic components. According to their Mixed Arithmetic-Perceptual Model, people interact with graphs to answer a certain question. They use search processes for indicators, encode the values of indicators, perform arithmetic operations based on these values, make spatial comparisons between the indicators, and finally respond to the question. The sequence of processing components may be different for each task. In the following, we argue that nonarithmetic and arithmetic processing components seem to influence students' graph reading performance.

Guthrie, Weber, and Kimmerly (1993) examined undergraduate students' understanding of graphs, tables, and illustrations. They found that two factors influenced their performance: i) an - what they called - elementary level questions factor reflecting students' ability to locate specific information in the respective graphs and tables and ii) an overall level questions factor referring to students' ability to perceive trends and patterns. The authors argued that perceiving trends and patterns requires an abstraction processes which should be independent from the processes of locating specific information in graphs (Guthrie et al., 1993). Later on, Åberg-Bengtsson (1999) analyzed the underlying dimensions of performance on the diagrams, tables, and maps items of the Swedish Scholastic Aptitude Test. Besides a general factor that influenced performance on all items, she found a quantitative factor as well as a complexity factor. The general factor reflected aspects of locating the necessary information in the respective graphs or tables (e.g., along the x and the y-axis of a line diagram, similar to the factor found by Guthrie et al., 1993) and thus nonarithmetic processing components that primarily include encoding. In contrast, the quantitative and the complex factor were more specific. The quantitative factor comprised all items that involved at least some calculations and thus arithmetic processing components. The third so-called complexity factor was associated with items that required multiple operations (numerical or not). The latter two factors may reflect students' ability to "mathematize" the graph reading. With mathematizing we refer to the students' ability to translate an ill-defined problem involving multiple steps into a mathematical structure (cf. Schoenfeld, 1989).

Additionally, there is evidence indicating that graph literacy is significantly associated with general cognitive abilities. For instance, Berg and Smith (1994) found graph literacy to be related to logical thinking and proportional reasoning in a sample of 7th, 9th, and 11th graders. Moreover, Padilla, McKenzie and Shaw (1986) observed that interpreting line graphs was associated with abstract-reasoning abilities in a sample of 119 7th, 9th, and 11th graders.

Furthermore, there is also evidence suggesting that there may be gender differences in graph literacy; boys were found to outperform girls (e.g., Lowrie & Diezmann, 2011; Åberg-Bengtsson, 1999; but see Curcio, 1987). Therefore, it is important to control for general cognitive abilities as well as gender when investigating the specific influence of basic numerical abilities on graph literacy.

In sum, we suggest that students are facing different challenges when solving graph reading tasks. First, the respective graph reading problem needs to be mathematized to enable them to – second – locate the relevant information, either numerical information or patterns and trends. The third challenge may be to correctly perform the necessary arithmetic operations.

In line with this idea, Curcio (1987) found that 7th graders' comprehension of graphs and tables was associated positively with their prior knowledge about graphical language as well as about the mathematical content necessary to solve the tasks. This is first evidence suggesting that (basic) numerical abilities may specifically contribute to graph literacy.

However, there are hitherto no empirical studies evaluating which specific basic numerical abilities influence graph reading performance. The current study pursued this question. Before describing the details of the present study, we will give a brief introduction to the idea of basic numerical abilities underlying numeric cognition.

II.1.2 Basic numerical abilities

Research has revealed that numerical cognition is not a unitary construct (e.g., Dowker, 2005). This means that numerical and mathematical skills build upon several basic numerical representations (e.g., Dehaene, 2009) and abilities, which are assumed to be the building blocks of numerical cognition in general and children's numerical development in particular (von Aster & Shalev, 2007). Basic numerical abilities include – but are not limited to – an understanding of non-symbolic and symbolic number magnitudes (e.g., Dehaene, Piazza, Pinel, & Cohen, 2003; Siegler, 2016) and a spatial representation of the corresponding magnitude (aka the mental number line, e.g., Booth & Siegler, 2008; see Fischer & Shaki, 2014 for a review), a verbal representation of number words, but also arithmetical facts (such as multiplication tables, e.g., Dehaene et al., 2003), a visual Arabic representation for understanding number symbols (e.g., Dehaene & Cohen, 1997; De Smedt, Noël, Gilmore, & Ansari, 2013 for a review), an understanding of the place-value structure of the Arabic number system (e.g., Moeller, Pixner, Zuber, Kaufmann, & Nuerk 2011; see Nuerk, Moeller, Klein, Willmes, & Fischer, 2011), as well as abilities on procedural and conceptual numerical knowledge (e.g., carry operation, but also commutative law).

A high level of mastery of these basic numerical abilities was repeatedly observed to be associated with actual numerical competencies but also predictive of future numerical competences as well as mathematical achievement in school (e.g., Moeller et al., 2011; Schneider, Grabner, & Paetsch, 2009).

In the following, a few examples will be given to illustrate how basic numerical abilities were found to influence later mathematical achievement. For instance, Kolkman, Kroesbergen, & Leseman (2013) observed that, already in kindergarten, children's early understanding of symbolic numbers as well as their skills in mapping these symbolic numbers to non-symbolic magnitudes, such as dot patterns, was an important predictor of their later numerical/mathematical development (see Schneider et al., 2017 for a meta-analysis). Furthermore, Booth and Siegler (2008) found that children's performance on locating the position of a target number on an empty number line not only correlated positively with children's actual basic numerical abilities (e.g., magnitude comparison, see also Link, Nuerk, & Moeller, 2014) but also predicted their ability to acquire new arithmetical competences in the future. Related to this, Moeller et al. (2011) found that early place-value understanding in first grade predicted later arithmetic performance as well as children's math grades two years later. Moreover, recent evidence suggests that basic numerical abilities are predictive of understanding more complex mathematical concepts taught in secondary school such as fractions. For instance, Bailey, Siegler, and Geary (2014) found that basic arithmetic abilities assessed in first grade predicted arithmetic performance on fractions in secondary school (see also Vukovic et al., 2014). Interestingly, however, to the best of our knowledge, there is currently no study investigating the association of basic numerical abilities and graph reading performance. This comes somewhat as a surprise given the above-mentioned importance of graph literacy in everyday life as well as in math curricula.

Therefore, the current study set out to evaluate the association of different basic numerical abilities with graph reading performance in a cross-sectional design. We hypothesized that basic numerical abilities implying the understanding of number magnitudes but also of basic arithmetic operations should be particularly relevant to graph reading. We expected such a pattern of results because graph reading tasks usually require a combination of the respective basic numerical abilities such as the identification and/or comparison of numerical magnitudes as well as (approximate) calculations.

II.2 Method

II.2.1 Sample

The total sample included $N = 812$ students recruited from German schools. We excluded 22 students older than 25 from the analysis. These students are likely to be in a higher developmental stage than the majority of students and may not be comparable to the majority of students because the average age of those was about 17 years. Furthermore, we excluded 15 students, because of missing demographic data, and 25 students because they did not complete the graph reading test. This left 750 students for the analyses. The final sample had a mean age of $M = 17.34$ years ($SD = 2.12$) and included 36% females.

II.2.2 Measures

II.2.2.1 Graph reading test

Graph reading was assessed by 14 items, seven of which addressed a bar and 7 a line graph. Respective responses were either a specific number or a specific label of the graph. The time for doing the whole test was restricted to 4 mins. Correctly answered items were considered in a sum score, which served as dependent variable.

Successful completion of the graph reading items required locating of specific information in the graph, several numerical operations such as magnitude comparisons, basic arithmetical operations, but there were also more complex modeling demands. An item was subject to modeling demands when it required an inference about which arithmetic operation had to be performed. For instance, to correctly answer a modeling item, participants had to infer that they would have to calculate the sum of two values.

Figure 6 gives an example of a bar graph. An exemplary item might have been "Which worker should be rewarded?". Students first needed to consider the context of the problem to infer that the worker who produced the highest quantity of shoes should be rewarded, (modeling demands), second, determine the worker with the longest bar graph (relational operation), third, sum up the two bars of potential candidates (addition) by reading off the exact values (localization of information), fourth, read the label to identify the worker who had produced the most shoes (localization of information). Please note that the item given in Figure 6 is for illustrative purposes only and was not an item actually used in the test.

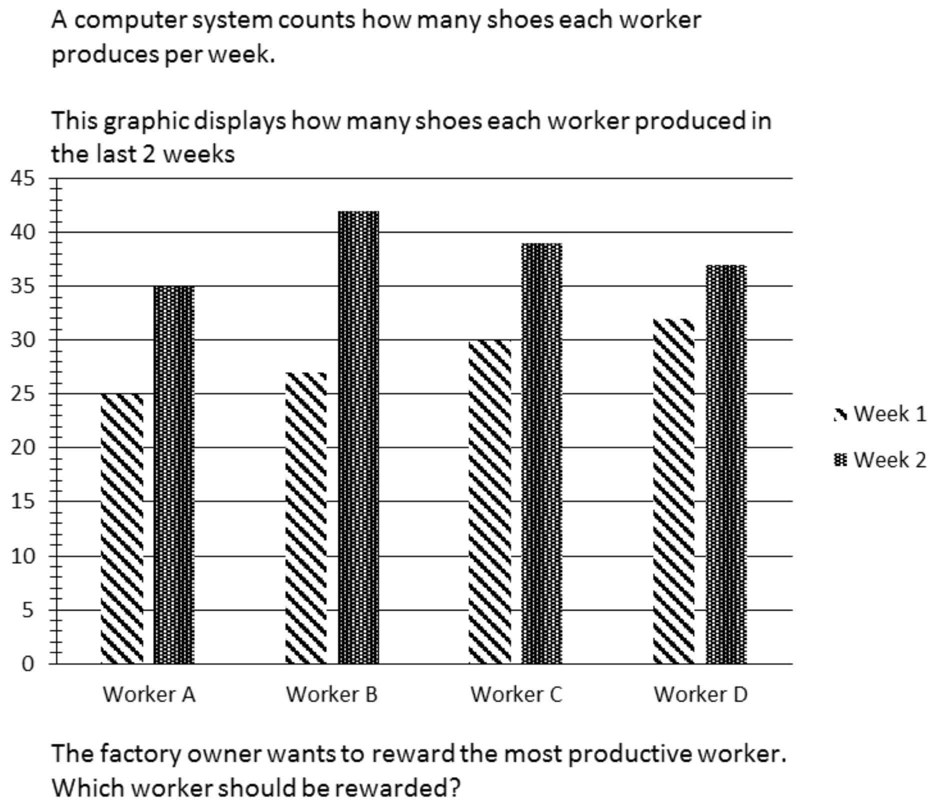


Figure 6. Schematic illustration of item format introducing item stem with a sample item in open answer format. Please note that this is not an item used in the experiment.

II.2.2.2 Basic numerical abilities

Different basic numerical abilities – ranging from non-symbolic magnitude comparisons to (approximate) arithmetic – were assessed by separate tests. All eight tests comprised only material which the students should have mastered already in primary school. Therefore, timing of all sub-tests was such to allow us concluding on the degree of automatization. For all tests, students were presented with examples familiarizing them with the topics. Unless indicated differently, correctly solved items were considered for sum scores which were used as dependent variable for all analyses. In the following, we will describe the tests used in more detail.

Non-symbolic magnitude comparisons: Students were presented with 16 pairs of dot clouds each and had to mark the cloud with the larger quantity of dots. Overall surface of the dot clouds was systematically varied in a way that for half of the items the surface with the lower quantity of dots was larger, which was reversed for the other half of the items. Time for this task was limited to avoid counting-based strategies (time limit: 1 minute).

Number line estimation: In the number line estimation task, students had to estimate the spatial location of a given number on a number line, of which only the endpoints were specified (e.g., the spatial location of 24 on a number line ranging from 0 to 100). Overall, the task comprised 24 items (6 items on a number line from 0 to 10 and 0 to 100, respectively, and 4 items each on a number line from 0 to 1.000, 0 to 10.000, and 0 to 100.000, respectively). Time for the task was limited to 1.5 minutes to ensure that students estimated the position of the target numbers intuitively. Mean absolute estimation accuracy in percent served as dependent variable and was used in the analyses. Items with no response were coded with an accuracy of zero.

Arithmetic operations: Arithmetic operations included: i) *Addition*, ii) *Subtraction* and iii) *Multiplication*. Each test comprised 36 arithmetic tasks in the order of ascending difficulty. Multiplications were restricted to single digit times single digit tasks, whereas addition and subtraction tasks covered the number range up to 10,000. Out of each operation, students had to solve as many tasks as they could within 2 minutes.

Approximate arithmetic: In the approximate arithmetic tasks, students had to choose the one out of two incorrect solution proposals that was closer to the correct result. The range comprised 16 addition and 16 subtraction tasks. For instance: "Which result is closer to 1546 - 687? Possible answers: 816 or 678". As in the basic arithmetic tests, difficulty level increased with every item. Students had 2 minutes to solve the tasks.

Basic geometry: Basic geometrical abilities were measured using 12 mirror image problems. Students were presented with a geometric shape (e.g., a rectangle) and an axis across which they had to mirror this shape by drawing the flipped form on the side opposite the mirror axis. For each correct line in a drawing, students were assigned one point. As a result, maximum scores for each item varied between 6 and 12. This resulted in a maximum score of 94 for this scale. A sum score was used as dependent variable.

Conceptual knowledge about arithmetic: In this task, students were presented with 40 pairs of arithmetic problems containing addition, subtraction, multiplication and division. Of each pair, one of the two tasks was already solved. Students had to decide whether the solution to the first problem helped them solve the second problem without having to calculate. For example: A) "Does 54 : 9 = 6 helps you solve 54 : 6 = ?" or B) Does 4 * 23 = 92 help you solve 92 : 4 = ?".

Additionally, we assessed children's g*eneral cognitive ability* as a covariate by two subtests of the German version of the Culture Fair Intelligence Scale 2-revision (i.e., sequence continuation, matrices; CFT 20-R; Weiß, Albinus, & Arzt, 2006). Subtests were administered as defined in the manual. In the sequence *continuation subtest*, students needed to find a logical continuation of a sequence of shapes. In the *matrix subtest*, students needed to do the same kind of conclusion for finding a logical shape for a blank cell of a matrix.

II.2.3 Procedure

All tests were administered during school hours in the students' classrooms by trained experimenters. Testing took maximally 90 minutes. Parents provided informed written consent; students were told that they could withdraw from the test at any time without negative consequences. Students above the age of 18 provided informed written consent by themselves; parents received information about the study. The study was approved by the local ethical committee as well as the regional school board.

II.2.4 Statistical Analyses

Prior to running the analyses, we checked for multicollinearity between variables. However, multicollinearity was no issue, because no two variables correlated higher than $r_{ij} > .8$ and no predictor variable showed a Variance Inflation Factor (VIF) > 10 (cf. O'Brien, 2007). Moreover, all variables except age were approximately normally distributed. The distribution of age was skewed, due to few students who were older than is usually expected for students in the investigated school types. Therefore, we log transformed the age variable prior to the analyses.

II.2.4.1 Multiple regression

We used multiple regression analysis to determine the significant predictors of graph reading performance with a False Discovery Rate (FDR) p-value adjustment for multiple testing (Benjamini & Hochberg, 1995).

II.2.4.2 Relative weight analysis

Additionally, we reported the relative weight of each predictor. The relative weight analysis (Johnson, 2000) addresses a problem caused by correlated predictors. Relative weight analysis uses a variable transformation approach to create a set of new predictors that are maximally related to the original predictors but are orthogonal to one another. In contrast to standardized regression weights, resulting relative weights represent the predictors' additive decomposition of the total model R^2. Two measures of relative weight can be calculated, the raw relative weight and rescaled relative weight. Raw relative weights add up to the R^2 of the model and the rescaled relative weights add up to 100%, representing the relative importance of a particular variable in regression model. Relative weights can be interpreted as the proportion of explained variance in criteria that can be appropriately attributed to each predictor variable (Tonidandel & LeBreton, 2015). Finally, we determined significance of relative weights using the procedure described by Tonidandel, LeBreton, and Johnson (2009).

II.2.4.3 Variables

For all analysis, we considered 29 predictor variables: the eight basic numerical abilities were assessed (i.e., addition, subtraction, multiplication, number line estimation, approximate arithmetic, conceptual knowledge, basic geometry, non-symbolic magnitude comparisons), as well as general cognitive ability, age, gender, and the interaction terms of age and gender with general cognitive ability and the eight basic numerical abilities.

We used an effect coding for gender (-1 = female, male = 1) and centered all continuous variables to be able to interpret the effect of interaction terms.

II.2.4.4 Statistical software

All statistical analyses were performed in the R environment (R Core Team, 2017). Multiple regression analysis was performed using the 'lm' function for fitting the linear models of the standard R package "state" (R Core Team, 2017). We used the 'p.adjust' function with the 'fdr' method to adjust *p*-values. Finally, we applied the syntax adopted from Tonidandel and LeBreton (2015) to conduct the relative weights analysis.

II.3 Results

The average graph reading score was $M = 6.71$ (SD = 2.28) out of 14 possible points. Therefore, graph reading showed sufficient variability. The predictor variables also showed sufficient variability (see Table 4), indicating that under the given time constraints none of the measures was either too easy or too difficult.

Table 4. Mean (M), Standard Deviation (SD) and obtained range of all measures.

	M	SD	Range
Graph reading	6.75	2.28	1 - 12
Addition	20.86	4.23	2 - 31
Subtraction	17.17	4.96	0 - 32
Multiplication	20.94	4.32	2 - 29
Number line estimation (%)	81.27	15.77	2.19 - 97.47
Approximate arithmetic	20.40	5.33	0 - 32
Conceptual knowledge	18.21	6.10	1 - 36
Basic geometry (%)	59.35	21.18	0.00 - 100.00
Non-sym. mag. comp.	18.75	3.13	1 - 24
G. cognitive ability	19.01	4.11	1 - 29

N = 750

The correlation matrix depicted in Table 5 indicated that almost all basic numerical abilities were significantly correlated with graph reading performance, as well as amongst each other. This held with only one exception: conceptual knowledge and non-symbolic magnitude comparisons were not related significantly. The control variables gender and age showed small correlations with some measures of basic numerical abilities (see Table 5). Gender was significantly correlated with graph reading performance. Age was negatively related to conceptual knowledge, basic geometry, and general cognitive ability. These significant correlations justified our decision to consider gender and age as control variables when evaluating the influence of basic numerical abilities on graph reading performance.

Table 5. Correlations between graph reading, basic numerical abilities, general cognitive ability, as well as gender and age.

Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. Graph reading	1										
2. Addition	.31**	1									
3. Subtraction	.39**	.68**	1								
4. Multiplication	.31**	.58**	.58**	1							
5. Number line estimation	.29**	.25**	.22**	.21**	1						
6. Approximate arithmetic	.29**	.40**	.44**	.36**	.37**	1					
7. Conceptual knowledge	.31**	.32**	.33**	.31**	.34**	.39**	1				
8. Basic geometry	.24**	.23**	.21**	.15**	.09**	.08*	.20**	1			
9. Non-sym. mag. comp.	.14**	.15**	.10*	.14**	.06*	.12**	.07	.08*	1		
10. G. cognitive ability	.45**	.34**	.35**	.32**	.17**	.21**	.28**	.41**	.24**	1	
11. Gender[a]	.10**	.11**	.23**	.14**	.24**	.21**	-.08*	-.06	-.06	.00	1
12. Log(age)	-.05	.06	.07	-.05	-.03	.05	-.12**	-.10**	.03	-.17**	.25**

Note: $^{**}p < .01$, $^{*}p < .05$. $N = 750$. [a]Code female = -1, male = 1

Multiple regression analysis. The linear multiple regression analysis including 29 predictors (i.e., basic numerical abilities, general cognitive ability, age, gender, and the interactions of basic numerical abilities with age and gender) explained about 34% of the variance [$R^2 = .34$, *adj.* $R^2 = .31$, $F(30,720) = 12.61$, $p < .001$] of graph reading performance. Four individual variables showed a significant effect on graph reading performance: general cognitive ability, number line estimation, subtraction, and conceptual knowledge (see Table 6). Inspection of the beta weights indicated that for all predictors better performance on the predictor was associated with better graph reading performance. General cognitive ability had a considerably larger effect on graph reading performance than conceptual knowledge, because the 95%- confidence intervals for the respective coefficients did not overlap. However, effect sizes of all other predictors were indistinguishable.

Additionally, we computed the relative weight of each variable (Johnson, 2000) to evaluate which predictors accounted for non-trivial variance of graph reading performance in contrast to regression weights that reflect incremental prediction. In case predictors are correlated they may not yield a significant incremental relationship.

Rescaled relative weights revealed that general cognitive ability accounted for a share of 28.54% of the explained variance in graph reading performance. Therefore, general cognitive ability seemed to be the best predictor for graph reading performance. Interestingly, the sum of basic numerical abilities accounted for a larger share of the explained variance in graph reading performance.

Furthermore, despite low regression weights consideration of rescaled relative weights indicated that addition (β > -.02), multiplication (β = .01*)*, approximate arithmetic (β = .03), and basic geometry (β =.05) explained a significant proportion of the predicted variance (i.e., between 5% and 6%). This showed that those may be relevant to the prediction of graph reading performance, but did not explain that there was enough incremental variance in the regression model to become significant. In contrast, gender and age had similarly low standardized regression weights (gender: β = .01; age: β = .01), but also very low rescaled relative weights (gender: *RS-RW* = 1.01; age: *RS-RW* = .42). Therefore, they may not be as relevant for the prediction of graph reading performance.

Table 6. Multiple regression results. Please note that for reasons of better readability only main effects are displayed (see appendix for a table including interaction terms).

	B	**β**	**[*L-CI,U-CI*]**	***RW***	***t***	***p***	***RS-RW* (%)**
Criteria = Graph reading performance [multiple R^2= .34, *adj*. R^2 = .31, $F(30,720) = 12.61$, p <.001]							
Intercept	6.71	.00	[-.06, .07]		80.53	.000	
G. cognitive ability	0.17	.30	[.22, .37]	.10	7.34	.000	28.54*
Subtraction	0.09	.20	[.11, .29]	.05	4.17	.000	13.86*
Number line estimation	2.13	.15	[.06, .19]	.04	4.07	.000	12.01*
Conceptual knowledge	0.04	.11	[.03, .18]	.03	2.86	.026	9.92*
Approximate arithmetic	0.01	.03	[-.04, .11]	.02	0.80	.748	6.12*
Multiplication	0.01	.01	[-.05, .11]	.02	0.34	.914	6.22*
Basic geometry	0.01	.05	[-.03, .11]	.02	1.27	.616	5.17*
Addition	-0.01	-.02	[-.14, .04]	.02	-0.35	.914	5.79*
Non-sym. mag. comp.	0.01	.01	[-.04, .09]	.01	0.29	.925	1.70
Gender[a]	0.07	.03	[-.04, .10]	.00	0.87	.748	1.16
Log(age)	0.14	.01	[-.06, .07]	.00	0.22	.954	0.46

Note: *b*: unstandardized regression weight, *β:* standardized regression weight, *L-CI:* lower boundary (2.5%), *L-CI:* upper boundary(97.5%), *RW:* raw relative weight (within rounding error raw weights will sum to R^2), *t:* t-value measures the size of the effect relative to the variation in sample data, *RS-RW:* relative weight rescaled as a, percentage of predicted variance in the criterion variable attributed to each predictor (within rounding error rescaled weights sum to 100 %). [a] code female = -1, male = 1.* significantly different from a random variable.

II.4 Discussion

In the current study, we aimed at investigating the influence of basic numerical abilities on graph reading performance. In two separate analyses, we first identified the significant predictors of graph reading performance and then evaluated the relative importance of individual basic numerical abilities. We observed general cognitive ability to be the most important predictor of graph reading performance. Beyond general cognitive ability, performance in number line estimation, subtraction and conceptual knowledge were significant predictors of graph reading performance. This finding is in line with the Mixed Arithmetic-Perceptual Model by Gillan and colleagues (Gillan & Lewis, 1994; Gillan, 2009) as the significant predictors are associated with arithmetical processes (subtraction and conceptual knowledge), and non-arithmetical processes (number line estimation and general cognitive ability). Furthermore, they also map well with Åberg-Bengtsson's (1999) differentiation between a general factor (reflected by general cognitive ability) and a quantitative factor (involving subtraction, number line estimation, and conceptual knowledge).

In the following, we will discuss these results in more detail by addressing each of the relevant predictors of graph reading performance one after the other. In particular, we will consider two questions: How may the respective predictor be related to graph reading (e.g., identifying specific information, relational operations, etc.) and which general cognitive processes and cognitive strategies may underlie its influence?

The most important predictor of graph reading was general cognitive ability. This influence of general cognitive ability is not surprising; however, there may be two reasons why general cognitive ability may have been of particularly predictive value for graph reading performance. First, graph reading problems may not be a common part of the daily routine in schools. General cognitive ability is often defined as the ability to solve new problems (e.g., Hartig, & Klieme, 2006). Therefore, general cognitive ability may be a significant predictor of graph reading performance. Second, the subtests of the Culture Fair Test used in the present study strongly rely on visual processing and analogical reasoning (Weiß et al., 2006). These processing components may also relate to the non-arithmetical processes, visual and visual imagery processes as were proposed to be relevant in graph reading by the Mixed Arithmetic-Perceptual Model by Gillan and Lewis (1994). Importantly, similar subcomponents of information processing have also been discussed to influence graph literacy by other authors (cf. Verschaffel, De Corte, & Lasure, 1994).

Moreover, number line estimation was observed to be the most important basic numerical ability to predict graph reading performance: Better number line estimation performance predicted better graph reading. Interestingly, number line estimation should relate to visual processes, such as performing spatial comparisons, value encoding and spatial differences as also reflected in the Mixed Arithmetic-Perceptual Model (Gillan & Lewis, 1994; Gillan, 2009). Therefore, the association between number line estimation and graph reading performace seems plausible because number line estimation also requires translation of symbolically and spatially-coded representations of quantities (i.e., to indicate the spatial location corresponding to an Arabic number). Furthermore, number line estimation is the only basic numerical ability – among those predictors considered significant for graph reading – that involves translations between symbolically and spatially-coded representations of quantity. This again is specifically required in graph reading. Moreover, localizing specific information in a spatial graphical set-up is a core component of graph literacy (Guthrie et al.,1993) which usually necessitates processing of quantitative information across different representational modalities. For instance, reading a specific value off a graph requires identifying the location of the respective information (e.g., the largest bar), and then to translate this spatial representation of magnitude into a symbolic representation by reading the corresponding position off a scale.

Additionally, recent research (e.g., Barth & Paladino, 2011; see Dackermann, Huber, Bahnmueller, Nuerk, & Moeller, 2015 for a discussion) has shown that number line estimation usually involves the application of so-called proportion-judgment strategies. This means that participants estimate the position of a target number by considering its relation to specific reference points (e.g., the middle of the depicted number line, i.e., 50 on a 0 to 100 number line). Very similar proportion-based strategies were observed to be applied for the estimation of bar graph lengths (Hollands & Spence, 1998). Also, most of the present graph reading tasks involved relational operations (i.e. using equal to, greater than, smaller than). Accordingly, students who performed well in number line estimation due to their ability to translate between symbolic and non-symbolic spatial representations of quantity and/or the application of efficient proportion-based strategies might also have performed better in graph reading tasks.

As another numerical predictor, subtraction was identified as a relevant and significant predictor of graph reading performance with better subtraction performance being associated with better graph reading. Most graph reading tasks in the present study involved arithmetic operations. However, subtraction was the only relevant and significant predictor of graph reading. There may be at least three reasons for that. First, variability was more pronounced for subtraction as compared to addition, which might account for the fact that subtraction and not addition was incorporated in the variable selection regression model. Second, addition is more prone to influences of arithmetic fact retrieval, which has been suggested for single-digit additions (e.g., Butterworth, Zorzi, Girelli, & Jonckheere, 2001; Pesenti, Thioux, Seron, & De Volder, 2000). In turn, the subtraction scale may better reflect students' procedural calculation skills. This argument is further corroborated by the fact that multiplication performance, which is widely agreed to reflect arithmetic fact retrieval (e.g., Butterworth et al., 2001; Pesenti et al., 2000), was not observed to be a significant predictor of graph reading. Finally, subtraction and addition performance were highly correlated. Considering the first two arguments, it is unlikely that addition explains unique variance over and above what had already been explained by subtraction.

Finally, students' conceptual knowledge was a significant predictor of their graph reading in a way that better performance in the conceptual knowledge scale was associated with better graph reading performance. In the conceptual knowledge scale, students needed to identify relationships between arithmetic operations rather than solve arithmetic problems by actually performing the necessary computations. As such, conceptual knowledge enables students to select and use problem solving strategies which are less resource demanding. For instance, in some tasks students had to compute means of two values depicted in a bar graph. The mean can be determined by reading the values off the two bars and then applying the arithmetic procedure [$M = (x_1 + x_2 + \ldots + x_n) / n$]. However, knowledge about relationships between arithmetic operations enables students to select a problem solving strategy demanding fewer resources. Such a strategy could be to take the spatial middle between the two respective graphs and then read off the respective value (Gillan, 1995). Therefore, higher conceptual knowledge may enable students to apply more efficient problem solving strategies in graph reading problems.

As indicated by the relative weight analysis multiplication, approximate arithmetic, and basic geometry accounted for a considerable amount of predicted variance of graph reading performance and might thus be relevant predictors of it as well. However, their influence turned out not to be significant in the multiple regression analysis – most probably because they share large parts of their variance with other predictors included in the model.

Multiplication may be a generally relevant predictor because of the influence of arithmetic fact retrieval (i.e., multiplication tables), which describes a different way of solving numerical tasks as compared to actual calculations. However, none of the problems in the graph reading test explicitly required multiplication. Moreover, approximate arithmetical abilities may be a relevant predictor because students with good approximation abilities may have advantages in judging the plausibility of possible results. However, all problems of the graph reading test required that students provided a specific numerical result, which could not be solved by approximation alone. Finally, basic geometry may be relevant for graph reading due to the necessity to process spatially represented information and may relate different pieces of spatial information explicitly to each other (e.g., to evaluate which worker produced more shoes by comparing the size of the bars, cf. Figure 1). All that is reflected in the basic geometry task, which also required participants to process spatial information and relate various pieces of it to produce the respective mirror images. Therefore, it may be necessary that these basic numerical abilities be acknowledged even though they were not identified as significant predictors of graph reading performance within the set of basic numerical abilities assessed in the current study.

So, what were the basic numerical predictors of graph reading? We identified number line estimation and subtraction performance as well as students' conceptual knowledge to be significant predictors of graph reading beyond the influence of general cognitive ability. Interestingly, these basic numerical abilities seem to be associated with graph reading because they reflect specific numerical processes and more general problem solving strategies required in graph reading. This indicates that graph reading performance may indeed be a mixture of arithmetical and non-arithmetical cognitive processes (Gillan & Lewis, 1994). Therefore, influences of tests involving computational skills such as the subtraction test seem obvious. Similarly, number line estimation implies processes such as the translation between symbolically and spatially-coded representations of quantities that are also necessary in graph reading tasks.

However, considering how number line estimation is achieved (by means of proportion-based strategies) and considering the observed influence of students' conceptual knowledge we conclude that the influence of basic numerical abilities goes beyond the level of providing mere numerical and arithmetical prerequisites for more applied everyday graph literacy. Instead, influence of number line estimation and conceptual knowledge indicates that mastery of basic numerical concepts allows for the application of more efficient problem solving strategies beyond mere arithmetical procedures.

II.4.1 Conclusions and Perspectives

As we pointed out above, we observed an influence of basic numerical abilities on graph literacy over and above influences of general cognitive ability. To the best of our knowledge, this is one of the first studies indicating that students who score high on basic numerical abiliites acquired in primary school, also perform better on more applied everyday graph reading tasks at a secondary school level. Additionally, basic numerical abilities and general cognitive ability which we identified as significant predictors of graph reading closely map with the distinction between arithmetical and non-arithmetical processes described by the Mixed Arithmetic-Perceptual Model of graph reading (Gillan & Lewis, 1994).

In sum, the present results underline the central importance of basic numerical abilities built up before and during the first years of formal schooling and later their impact on more complex math skills, which, in turn, are relevant for educational achievement and everyday life. We are faced with different forms of graphs in various situations every day (Åberg-Bengtsson, 2006), for instance regarding results of elections, stock market development, labor market information, and so on. As indicated by the results of this study, the foundation for graph literacy is laid in primary school. However, in secondary education, teachers usually expect students to have acquired a sufficient level of basic numerical abilities. Therefore, teachers do not focus on them later again, which is most obvious with regard to basic arithmetic operations. As a consequence, students with deficits in these areas may experience severe problems in solving more complex tasks such as graph reading. Even though the present results rely on cross-sectional data, they indicate that even in secondary education, it seems worth reviewing elementary school knowledge in the context of real-world math problems. It would be desirable for future research to clarify the role of basic numerical abilities on graph reading in a longitudinal study to help understand the developmental trajectory of their association with graph reading.

Chapter III. Interpreting process measures in text-graphics comprehension

Process measures such as gaze patterns have been successfully used to investigate cognitive processes in educational research. However, process measures say nothing about the success or failure of the underlying cognitive processes, and process measures can have fundamentally different interpretations. In text-graphics comprehension, transitions between text and graphics can either be interpreted as integration, i.e., the building of referential connections, or as disorientation, i.e., the inability to identify relevant information. In this study, we argue that different interpretations apply depend on whether processing is more controlled or more automatic. Consequently, different interpretations apply during the (more controlled) initial reading of task material and the (more automatic) completion of tasks. Furthermore, prior knowledge as well as reading and graph comprehension abilities should influence individuals' ability to process material automatically. The results of two studies demonstrate that taking more time and performing more text-graph transitions is positively associated with task success during initial reading, but negatively associated with task success during task completion. Subsequently, we found that prior knowledge moderates the effect of time taken during initial reading and task completion. Our results indicate that the interpretation of process measures depends on comprehension phase (initial reading vs. task completion) and degree of prior knowledge. Furthermore, we discuss theoretical implications for multimedia learning and educational research methods.

III.1 Introduction

Eye movements have been successfully used in educational research to investigate cognitive processes in learning and testing situations (Scheiter & Eitel, 2017). Eye-tracking provides valuable process information. However, it tells us nothing about the success or failure of comprehension processes (Hyönä, 2010). The interpretation of process measures during text-graphic comprehension can be ambiguous. For instance, many transitions between text and graphics can indicate either text-graphic integration (engaging in global coherence formation) or disorientation (a lack of ability to find relevant information). In this paper, we address this ambiguity by analyzing the relationship between transitions between text and graphs and time-on-task with comprehension outcomes.

First, we describe cognitive processes essential to text-graphic comprehension and process measures that can be used to assess these cognitive processes. We operationalize cognitive processes of text-graphic comprehension with local and global coherence formation and process measures with time-on-task and text-graphic transitions. We address a specific type of graphics, namely, graphs. Graphs are quantitative axis diagrams that use an apposed-position language to represent relationships between data points (Lowrie, Diezmann, & Logan, 2012) and see text-graph comprehension as a particular case of text-graphic comprehension. Second, we discuss how process measures are indicative of cognitive processes during the initial reading of text-graph material and the completion of comprehension tasks. In two studies, we analyze the association between task success, time-on-task, and text-graph transitions during initial reading and task completion. In the second study, we additionally investigate whether prior knowledge as well as reading and graph comprehension ability moderate the effect of process measures on task success.

III.2 Theory

III.2.1 Text-graph comprehension

Combinations of text and graphics are a major design feature in science textbooks (e.g., Schnotz, 2005), where they serve as a useful tool for enhancing learning outcomes (e.g., Butcher, 2014; Carney & Levin, 2002; Mayer, 2005). However, students can only benefit from illustrated texts if they can mentally integrate information from the text and graphics (Seufert, 2003). Integration can be particularly challenging when the graphics are quantitative axis diagrams (Ullrich et al., 2012), to which we refer to as graphs. There is empirical evidence from both online (i.e., eye movement data; e.g., Johnson & Mayer, 2012; Mason, Pluchino, Tornatora, & Ariasi, 2013) and offline indicators (i.e., cross-modal memory intrusions, e.g., Schüler, Arndt, & Scheiter, 2015; Arndt, Schüler, & Scheiter, 2015) that integration is essential to the comprehension of illustrated texts.

Text-graph comprehension is successful when students are able to identify relevant information from each representation format, organize it into coherent modality-specific mental models (form local coherence, Seufert, 2003), and identify correspondences between text and graphs (form global coherence, Seufert, 2003).

Local coherence describes a person's understanding of each of the given representations (Ainsworth, Bibby, & Wood, 2002), or in other words the selection and organization of information into modality-specific mental models (Mayer, 2005). Local text coherence enables students to distinguish between the surface and deep structures of sentences and texts. The deep structure of a sentence is a theoretical construct which makes the underlying logical and semantic relations explicit and is independent of a specific sentence with specific syntax and specific words (Chomsky & Halle, 1965). Accordingly, local text coherence enables a person to match sentences with the same deep structure despite their different surface structures (Royer, Hastings, & Hook, 1979). This process requires a mixture of linguistic knowledge, word knowledge, and reading skills (Kintsch, 1988). Individuals who are fluent in reading comprehension take less time to select and organize relevant information from the text.

Local graph coherence enables students to distinguish between the surface and deep structures of graphs. In this paper, we refer to a specific type of graphics, namely, graphs. Graphs are quantitative axis diagrams that use an apposed-position language to represent relationships between data points (Lowrie, Diezmann, & Logan, 2012). Graphs use two-dimensional space to visualize relationships and convey meaning. The deep structure of a graph makes the conceptual and logical relationships between variables explicit, independent of the specific array of points and visual features such as dots, lines, and areas (Schnotz & Baadte, 2015). Understanding a graph requires grasping the meaning of the graph by constructing a mental model of its content (Kosslyn, 1989; Pinker, 1990). In contrast to pictures and schematic drawings, for instance, understanding the content of graphs requires knowledge about graphical conventions (Lowrie et al., 2012; Shah & Freedman, 2011). Individuals who are fluent in graph comprehension may take less time to select and organize relevant information from the graph.

Global processes link texts and graphs at a conceptual level (Ainsworth, 2006); in other words, they integrate modality-specific mental models (Mayer, 2005) or integrate a proportional representation of text with a mental model of its content (Schnotz, 2005). Global coherence enables students to map the deep structures of text and graphs onto one another. According to Schnotz et al. (2014), prior knowledge organizes the formation of global coherence. Therefore, individuals' prior knowledge should influence global coherence formation.

In this study, we investigate how students achieve global coherence. In the next section, we discuss how we elected to quantify global coherence formation.

III.2.2 Process measures in text-graph comprehension

We define process measures in opposition to outcome measures. Process measures capture behavior regulated by the individual, which is associated with and temporarily upstream to the outcome. Process measures can be indicators of cognitive processes. However, the two are not identical. In this paper, we address two common process measures: time-on-task and eye movements (Scheiter & Eitel, 2017). Below, we discuss the effects of time-on-task and transitions between text and graphs on task success.

III.2.2.1 The relationship between time-on-task and task success.

Both positive and negative influences of time-on-task have been investigated in the context of skill assessments (Goldhammer et al., 2014; Naumann, & Goldhammer, 2017). There are two ways to explain the relation between time-on-task and task success in skill assessments. Taking more time to work on a task may be positively related to task success as the task is completed more thoroughly and answers are more *elaborate*. On the other hand, the relationship may be negative if working faster and more *fluently* reflects a higher skill level. Goldhammer et al. (2014) found a negative relationship between time-on-task and task success. They argue that the automatic nature of reading processes at the word, sentence, and local coherence levels leads to a negative time-on-task effect. The effect is negative because both faster and more accurate performance are associated with reading skills such as phonological recoding, orthographic comparison, or the retrieval of word meanings from long-term memory (Richter, Isberner, Naumann, & Neeb, 2013). More prior knowledge should also allow students to perform faster and more accurately. However, text-graph comprehension does not only require reading skills up to the local coherence level. According to Goldhammer et al., when reading becomes more controlled, the relationship between time-on-task and comprehension outcome can become less negative or even positive. Text-graphic comprehension requires a higher level of controlled processing to establish global coherence by making referential connections between text and graphics (Mason et al., 2013; Mason, Tornatora, & Pluchino, 2015).

In sum, time-on-task has a negative effect on task success when tasks require automatic processing. If the material allows for automatic processing, individuals with greater fluency are likely to be faster and more accurate at the same time, while individuals who are less fluent should be slower and make more mistakes. Time-on-task has a positive relationship with task success when tasks require more controlled processes (e.g., rereading, building referential connections). If the task material is complex, even highly skilled individuals cannot be fast and accurate at the same time.

Consequently, text-graph comprehension involving only local text coherence formation may be a mostly automatic process, while building referential connections between text and graphics may be a more controlled process.

III.2.2.2 The relationship between text-graph transitions and task success.

Multimedia learning research has investigated the transitions between text and graphs using various methods (e.g., Hegarty & Just, 1993; Johnson & Mayer, 2012; Ozcelik, Karakus, Kursun, & Cagiltay, 2009). Similarly to time-on-task, there are two ways of explaining the relationship between text-graphic transitions and task success. First, referential connections allow students to *integrate* text and graphics mentally (Seufert, 2003). More referential connections between text and graphs facilitate comprehension and learning. Consequently, comprehension is positively related to transitions between text and graphics. For instance, more transitions between text and graphics during second-pass reading have been found to have a positive influence on outcome measures, such as verbal and graphical recall and transfer of knowledge (Mason et al., 2015, 2013). Transitions between problem statements and graphs are positively related to the probability of solving the problem (Ögren, Nyström, & Jarodzka, 2017). The authors argue that these transitions reflect controlled processes. Furthermore, some multimedia studies have shown that design features that increase text-graphic transitions also increase recall and transfer performance (spatial contiguity: Johnson & Mayer, 2012; signaling: Ozcelik et al., 2009).

On the other hand, the relationship between text-graph transitions may be negative because it reflects a lack of ability to make referential connections between text and graphs. Students may be *disoriented*, alternating rapidly between text and graph while failing to complete the task correctly. Schwonke, Berthold, and Renkl (2009) found a negative or zero correlation between conceptual understanding and text-diagram, text-equation and diagram-equation transitions in a multiple representations study.

In summary, both positive and negative associations between comprehension outcomes, time-on-task, and text-graph transitions are plausible because different mechanisms play a role (See Table 7). First, the construction of a globally coherent mental model (which can be achieved via elaboration and integration) increases the chance of an accurate answer; second, a smooth, undisrupted answering process indicates that the answer will be correct.

Previous research shows that various task and individual characteristics influence the extent and direction of the association between process measures and task success. In the following section, we introduce the idea that this association can depend on the comprehension phase.

Table 7. Which Underlying Cognitive Processes May Be Inferred from the Positive and Negative Associations between Process Measures and Comprehension Outcome.

		Time-on-task	**Text-graph transition**
Association with comprehension outcome	positive	Elaboration (e.g., Goldhammer et al., 2014)	Integration (e.g., Mason et al., 2013; Mason et al, 2015)
	negative	Fluency (e.g., Goldhammer et al., 2014)	Disorientation (e.g., Schwonke et al.,2009)

III.2.3 Initial reading and task completion

The degree to which processing is automatic and controlled not only depends on the material's level of complexity and the individual's skill level (Goldhammer et al., 2014), it may also depend on the comprehension phase. Schnotz et al., (2014) distinguished between two phases of text-picture comprehension: initial reading and task completion. In the context of solving multiple-choice questions and complex problem-solving, authors distinguish between knowledge acquisition and knowledge application (Greiff, Wüstenberg, Molnár, Fischer, Funke, & Csapó, 2013; Lindner, Eitel, Strobel, & Köller, 2017).

Individuals follow different processing goals during these phases. The aim of the initial reading phase is to construct a coherent mental model of the content (Schnotz et al. 2014; Schnotz & Wagner, 2018). Mental model construction is a controlled process. Therefore, it is plausible that investing more time into mental model construction will result in a more *elaborated* understanding of the material. In contrast, the task completion phase does not aim for mental model construction per se, but for the quick identification of information relevant to the task solution. Individuals who process more *fluently* identify task-relevant information within the text and graph faster and are more likely to be correct. Individuals who take longer may not have been able to find definite solution, and are forced to ruminate and finally guess the answer. Due to the fundamentally different goals of the two comprehension phases, we hypothesize that the relationship between processing measures and comprehension outcomes differs depending on phase.

III.2.4 Hypotheses

Our hypotheses are specified on the basis of two assumptions: (1) Time-on-task and text-graph transitions are measures of cognitive processing (eye-mind hypotheses: Carpenter & Just, 1975) und (2) individuals followed the instructions.

- Process measures during initial reading and task completion affect task success in opposing directions (H1)
    - Time-on-task and text-graph transitions during initial reading are positively related to task success due to *elaboration* and *integration* processes, respectively (H1.1).
    - Time-on-task and text-graph transitions during task completion are negatively related to task success because individuals with good comprehension of the material are more *fluent* and quickly identify task-relevant information (H1.2).
- Finally, global coherence is the critical cognitive process. Therefore, text-graph transitions explains task success above and beyond time-on-task. A model including both time-on-task and text-graph transitions has better fit than a model with time-on-task alone (H2).

III.3 General method

III.3.1 Material

III.3.1.1 Initial reading

For initial reading, we chose three different topics from biology, namely population dynamics, action potentials, and sleep cycles. Figure 7 shows the respective explanatory graphs from the initial reading material.

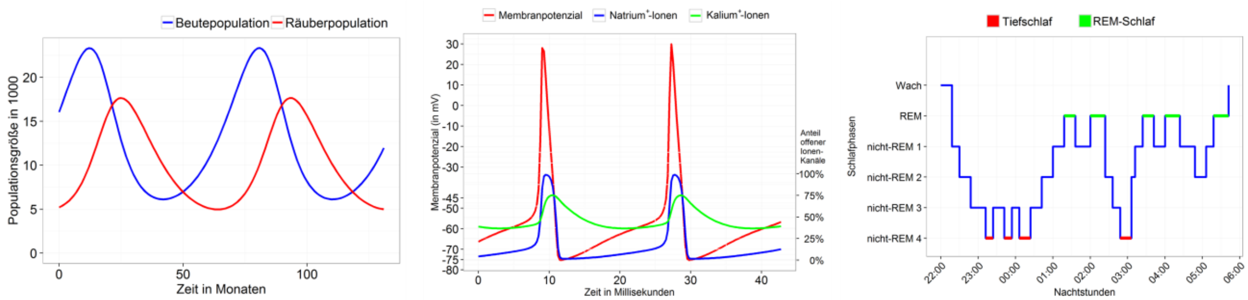
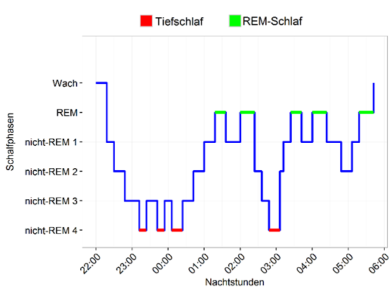


Figure 7. Example graph for each topic: population dynamics (left), action potentials (middle), and sleep cycles (right).

The three topics included text with 200 to 217 words. Each text consisted of 14 sentences and two paragraphs. The texts provided an overview of the interaction between predator and prey populations, the triggering of action potentials in neurons, and the sequence of sleep cycles. Figure 8 shows the initial reading text for the sleep cycle topic.

Are we sleeping deeply all night?

The activity of our brain can be measured during our sleep based on brain waves. These brain waves show that sleep is not a continuous state. For healthy people, different phases of sleep can be distinguished: non-REM sleep phases and REM sleep phases. REM stands for "Rapid Eye Movement". REM sleep is distinguished by particularly heavy muscular and physiological activity and is most similar to the wake state from all sleep phases. Non-REM sleep can in turn be divided into four phases. In the first non-REM sleep phase, sleep is still very easy and becomes ever deeper until the fourth non-REM sleep phase is reached.



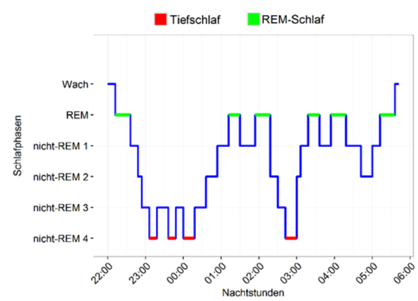
Sleep begins with the first non-REM sleep phase and then becomes ever deeper to the fourth non-REM sleep phase. The fourth non-REM sleep phase is therefore also referred to as deep sleep. The sequence from the first non-REM sleep phase to the REM sleep phase is referred to as sleep cycle. REM sleep terminates each cycle. A sleep cycle takes about one and a half hours, with a healthy person going through 4 to 6 sleep cycles in one night. The sequence of the sleeping phases and their relative shares within the sleep cycle change over the course of the night. The deep non-REM sleep phases (3, 4) are predominant in the first third of the night, after which the light (1,2) non-REM and REM sleep phases are predominant. As a rule, the first two sleep cycles even include all low sleep phases of the night.

Figure 8. Initial reading page for the sleep cycle domain (translated into English).

III.3.1.2 Task completion

After the initial reading phase, participants answered text-graph integration items for each topic. The texts included two different task types. The first task type required participants to find a contradiction between text and graph (see Figure 9). The second task type required participants to select graphs that matched the text. Hence, solving an item required integrated comprehension of the text and graph. However, the information presented in the item material was sufficient to identify the correct answer. Comprehension of the initial reading material should be helpful. However, a good understanding of the material was not sufficient to answer the items, because each item included new graphs.

Which sentance contradicts this graphic?



The activity of our brain can be measured during our sleep based on brain waves. These Brainwaves show that sleep is not a continuous state. For healthy people, different phases of sleep can be distinguished: the non-REM sleep phases and the REM sleep phases. REM stands for "Repeated Eye Movement". The REM sleep is distinguished by particularly great muscular and physiological activity and is most similar to the wake state from all sleep phases. The non-REM sleep can in turn be divided into four phases. In the first non-REM sleep phase, sleep is still very easy and becomes lower and lower over phases 2, 3 and 4.

Sleep begins with the first non-REM sleep phase and then becomes ever deeper to the fourth non-REM sleep phase. The fourth non-REM sleep phase is therefore also referred to as deep sleep. The sequence from the first non-REM sleep phase to the REM sleep phase is referred to as sleep cycle. REM sleep terminates each cycle. A sleep cycle takes about one and a half hours, with a healthy person going through 4 to 6 sleep cycles in one night. The sequence of the sleeping phases and their relative share in the respective sleep cycle change during the night. The deep non-REM sleep phases (3, 4) are predominate in the first third of the night, after which the light (1,2) non-REM and REM sleep phases are predominantly. As a rule, the first two sleep cycles even include all low sleep phases of the night.

The graphic does not contradict the text.

Figure 9. Sample item on sleep cycles. The item is solved by identifying the contradiction between graph and text. The text states that people enter sleep in the first non-REM phase. However, the graph depicts entering sleep via an REM phase. Participants have to click on the first sentence of the second paragraph to answer the item correctly.

III.3.2 Procedure

After filling out the informed consent form, participants worked through the items for all three topics in a random order. They began each topic by initially reading the task material. Participants were instructed to carefully process the material for general comprehension. The initial reading time was set to a minimum of one minute to prevent participants from rushing through the material. Afterwards, participants proceeded to the four text-graph comprehension items in the task completion phase. A simple example for each task type was presented to demonstrate how each task worked. The example item had to be answered correctly for participants to be able to proceed.

The items were then presented in random order; however, items of the same type were presented in sequence to reduce confusion due to task switching. This procedure was repeated for each of the three domains. Participants finished the study by answering a short demographic questionnaire. The study took 25 minutes on average.

III.3.3 Process measures

III.3.3.1 Equipment.

The study was performed with an HP ZBook 15. Stimuli were presented with Experiment Center (v. 3.0.128) in the web browser Firefox on an HP 15.6 inch computer screen with a resolution of 1920 x 1080 pixels and a refresh rate of 260 Hz. Eye movements were recorded at 260 Hz with the SMI eye tracker from SensoMotoric Instruments running iView X (v. 2.7.13).

III.3.3.2 Areas of interest.

For the text-graph comprehension items, we divided the initial reading material and text-graph items into two areas of interest: the text area and graph area. Due to the item design, we separated the text area and graph area with a divide along the Y-axis.

III.3.3.3 Data preparation.

Our analyses use raw time-on-task in minutes and the number of text-graph transitions (i.e. Saccade count; Lai, Tsai, Yang, Hsu, Liu, Lee, & Tsai, 2013). We ensured data quality by inspecting every trial for plausibility, eventual drift and by-trial tracking rate. Details on the data preparation procedure can be found in the appendix.

III.3.3.4 Treating missing values.

We considered the eye-tracking data from 68 trials as not trustworthy on the basis of a plausibility check and tracking rate. We removed three individuals from the sample because they had fewer than four valid trials. The remaining 38 (10.50%) non-trustworthy trials were spread across individuals and items. We avoided listwise deletion because it reduces the testing power and can potentially result in biased estimates. Since the responses and response times still provided reliable information about the trials, we estimated effects using multiple imputation procedures (Buuren, & Groothuis-Oudshoorn, 2011). More specifically, we applied predictive mean matching (PMM) because it produces unbiased imputations even when variables are not normally distributed (Rubin & Schenker, 1986). In principle, PMM constructs a metric for matching cases with missing values to similar non-missing cases within the same data set. Cases were selected on the basis of individual, item, domain, response accuracy, and response time. We matched each missing value with the five non-missing values that had the closest predicted values, following the recommendation for smaller data sets (Morris, White, & Royston, 2014). The variables were imputed in a particular order. We started with the variables containing the most missing values and ended with the variable with the fewest missing values. Plots indicated convergence after a few iterations (i.e. 10 iterations).

III.3.4 Statistical analyses

III.3.4.1 Process measures effects.

We used the generalized linear mixed model (GLMM) framework (e.g., Baayen, Davidson, & Bates, 2008; De Boeck et al., 2011) to investigate the role of time-on-task (time) and text-graph transitions (transitions), on the probability of answering a text-graph comprehension item correctly. A generalized linear mixed model is a linear regression model that includes fixed and random effects (b) using a logit link function ($\eta_{pi} = \ln(\pi_{pi} /(1 + \pi_{pi}))$) to transform a continuous linear component η_{pi} into the probability of obtaining a correct response π_{pi}. In our analysis, we define both (fixed) effects that are constant across items and persons and (random) effects that vary across items and persons (cf. De Boek). In other words, we first specify a regression model that assumes that the probability of an item being answered correctly depends on the ability of the individual and the difficulty of the item. The probability of an item being answered correctly increases with the individual's ability, while the probability of an item being answered correctly decreases with the difficulty of the item. In addition to the random item effect, we include task characteristics as fixed effects. We consider this the baseline model ($\pi_{pi} \sim \eta_{pi} = \beta_0$ (grand mean) + (individual skill b_{op}) + (relative easiness b_{0i}) + $\beta_{1\text{-}3}$ (tasks characteristics) + r_{pi}).

We include the domain of the item using dummy coding with two factors, one for action potentials and one for sleep cycles. In our study, all items were drawn from the domains of either population dynamics (action potential: 0, sleep cycle: 0), action potentials (action potential: 1, sleep cycle: 0) or sleep cycles (action potential: 0, sleep cycle: 1). We also include task type as a dummy-coded factor. The task type is either mapping from text to graph (text to graph: 0) or mapping from graph to text (text to graph: 1). Therefore, the intercept must be interpreted as the probability of an answer being correct when submitted by an average person for an average item from the domain of populations dynamics and the task type mapping from graph to text. However, since we are primarily interested in the effect of the process measures, task characteristics serve as control variables.

We then add the process measures to the model as fixed effects to investigate how they influence the probability of correct responses. In sum, we predict the probability of a correct response based on the person's ability, the difficulty of the item, the time-on-task, and the number of text-graph transitions or focus on the graph, while controlling for item characteristics. The effects we report are log odds ratios.

III.3.4.2 Estimates for multiply imputed data.

The multiple imputation resulted in five different imputed and therefore complete data sets. The estimates we report in the following analysis are averages of these five complete data sets. The rules we applied for combining the separate estimates, standard errors, confidence intervals and p-values are based on Rubin & Schenker (1986). The number of degrees of freedom was calculated using the method by Barnard and Rubin (1999, as cited by Buuren, & Groothuis-Oudshoorn, 2011).

III.3.4.3 Statistical software.

We applied the glmer function of the R package lme4 (Bates, Mächler, & Bolker, 2014), to estimate the presented GLMMs. We used the mice function of the R package mice (Buuren, & Groothuis-Oudshoorn, 2011) as well as miceadds (Robitzsch, Grund, & Henke, 2017) to perform the multiple imputations. The 'pool' function was used to average estimates across imputed data sets. The R environment (R Development Core Team, 2012) was also used to conduct logistic regression analyses.

III.4 Study 1

In the first reported study, we recorded students' eye movements during the initial reading of 3 text-graph combinations on biological topics and completion of 12 text-graph comprehension items. Our goal was to investigate the association between text-graph transitions, time-on-task, and task success.

III.4.1 Method 1

III.4.1.1 Sample

34 students (23 female; *M* = 20.46 years, *SD* = 3.36) from a university in southern Germany participated voluntarily in the study for course credit. Participants were psychology or cognitive science students. Data from 5 participants had to be excluded from data analysis because the data quality was not sufficient. More detailed data preparation criteria was reported in the method section. Participants were invited to the eye-tracking laboratory in groups of up to 10 individuals.

III.4.2 Results 1

III.4.2.1 Descriptive

The overall average initial reading time was 96.32 seconds and varied slightly across the different materials (Appendix Table 2 provides detailed statistics on the differences between the population dynamics, action potentials, and sleep cycles material). The overall average number of text-graph transitions during the initial reading phase was approximately 11. Participants tended to perform fewer transitions while reading the action potential material.

Overall, 53% of all responses were correct (Appendix Table 3 provides detailed statistics on task completion phase by item). The accuracy rate varied across items and ranged from 13% to 71%. Item 6 was answered correctly in 13% of all cases. The rest of the items varied between 40% and 71%. Responses took 52.51 seconds on average. The time-on-task ranged from 38.72 seconds to 66.81 seconds. Overall, about 11.56 transitions between text and graph were performed. The items of the second task type had a higher accuracy rate and required fewer text-graph transitions.

In summary, average accuracy rate, time-on-task, and text-graph transitions varied both across items and across persons. We assume that the accuracy rates and means of the process measures varied randomly across items. Additionally, we account for possible systematic influences by including the topic and task type in the analysis.

III.4.2.2 Item and person effects.

We analyze the effect of item and person on the probability of correct answers. The overall accuracy rate of 53.63% corresponds to a ‘grand’ intercept estimation of $\beta_{intercept} = 0.14$; $z = 0.57$; $p = .572$. The $\beta_{intercept} = 0.14$ estimate is the marginal log-odds of a correct answer for an average person completing an average item. The *p*-value indicates that the intercept is not significantly different from 50%.

The average probability of correct answers is 54%. However, another critical aspect is the range of probabilities of a correct answer across items and persons. The effect of process measures is difficult to detect if the probability of a correct answer is already very high (or low). For instance, a person with very high ability working on a very easy item cannot improve his or her probability of a correct answer by taking more time. However, in this study, the probability of a correct answer for the person with the highest ability working on the easiest item and the person with the lowest ability working on the most difficult item ranged from 88.02% to 11.57%, respectively. This range indicates that we are not investigating extreme combinations of ability and difficulty, and that process measures could still potentially have positive or negative effects on the probability of success across all responses.

Item characteristics. The significant intercept means that the probability of a correct response for the reference items (item characteristics: text-to-graph and population dynamics) is higher than 50%. More specifically, the probability of a correct response for the reference category is 76.20%. Therefore, the intercept represents the least difficult combination of item characteristics. In contrast, the most difficult combination of item characteristics is the topic of sleep cycles and the first task type, where the probability is 39.03%.

III.4.2.3 Process measures and the probability of correct answers.

Time-on-task. Time-on-task does not have a significant effect on task success during initial reading ($\beta_{ToT:IR} = 0.57$, $SE = 0.31$, $z = 1.86$, $p = .063$). Time-on-task has a negative effect during task completion ($\beta_{ToT:TC} = -0.93$, $SE = 0.23$, $z = -4.07$, $p < .001$). However, the effect of time-on-task is different during initial reading and task completion, since the lower boundary of time-on-task during initial reading ($\beta_{ToT:IR\ CIlow} = -.031$) does not overlap with the upper boundary of time-on-task during task completion ($\beta_{ToT:TC\ CIupper} = -0.48$).

The effect of time-on-task during task completion means that the probability of a correct response decreases (from the intercept and an average person) by 55.82% when the item is worked on for one minute longer. This result supports the hypothesis that the effect of time-on-task during initial reading and task completion is different, since the effect during initial reading is positive in tendency, and the effect during task completion is negative. The negative effect of time-on-task during task completion is in line with the hypothesis that the quick identification of task-relevant information leads to correct responses (H3). The positive yet insignificant tendency during initial reading at least does not directly contradict our hypothesis that taking more time during this phase reflects more elaborate and thorough processing of the material.

Text-graph transitions. Text-graph transitions have no significant effect on task success during initial reading ($\beta_{TGT:IR} = 0.02$, $SE = .02$, $z = 0.88$, $p = .327$), but they do during task completion ($\beta_{TGT:TC} = -0.05$, $SE = 0.02$, $z = -2.49$, $p < .001$). Again, even though the effect of transitions during initial reading is not significantly different from zero, text-graph transitions have a significantly different effect during initial reading than they have during task completion. Their confidence intervals do not overlap ($\beta_{TGT:IR\ CIlow} = -0.011$; $\beta_{TGT:TC\ CIupper} = -0.024$). The effect of text-graph transition during task completion is equivalent to a roughly one percent (0.91%) decrease in the probability of a correct answer (assuming an intercept item and an average person) when one more transition occurs. These results support our expectation that text-graph transitions during task completion indicate disorientation and the inability to find task-relevant information. The positive yet insignificant tendency for more text-graph transitions during initial reading again at least does not directly contradict our interpretation that transitions are indicative of integration processes and global coherence formation. The effect of text-graph transitions during initial reading may be particularly important because text-graph transitions represent eye movement behavior during a much shorter period.

Table 8. Study 1: Generalized Mixed Effect Regression with Person and Item as Random Effects, and Time-on-Task (ToT), Text-Graph Transitions (TGT) during Initial Reading (IR) and Task Completion (TC) as Fixed Effects.

		Baseline				**ToT**				**TGT**			
		est.	SE	*z*	*p*	est.	SE	*z*	*p*	est.	SE	*z*	*p*
Fixed Effects													
ToT	IR					0.57	.31	1.86	.063				
	TC					-0.93	.23	-4.07	.000				
TGT	IR									0.02	.02	0.88	.327
	TC									-0.05	.02	-3.49	.001
Item char.	Intercept	1.09	.38	2.85	.005	1.07	.64	1.67	.096	1.20	.42	2.87	.005
	Task ty.	-0.83	.32	-2.63	.009	-0.97	.33	-2.97	.003	-0.36	.32	-1.13	.261
	Action p.	-0.53	.35	-1.46	.146	-0.40	.36	-1.10	.272	-0.38	.33	-1.22	.222
	Sleep cy.	-0.68	.36	-1.88	.061	-0.74	.37	-2.00	.047	-0.68	.33	-2.06	.041
Random Effects													
	$\tau_{00,person}$	.73				.88				.78			
	$\tau_{00,item}$	.09				.08				.03			
	N_{id}					29							
	N_{item}					12							
	Obser.					348							
	Missings	0				ToT:IR=0; ToT:TC=2				TGT:IR =10; TGT:TC=38			
	Deviance	467.2				367.4				379.3			

Note: est = log-odds ratio estimates, SE = Standard Error

III.4.2.4 Incremental fit.

With the combined model, we tested whether text-graph transitions have added predictive value for determining task success. The combination of chi-square statistics for the five datasets of multiply imputed data shows a non-significant result $\chi^2(2) = 0.31$, p = .731. Thus, for this sample, we cannot distinguish between the effects of text-graph transitions and time-on-task due to the high correlation between the two.

III.4.2.5 Process measures and random effects.

The bottom of Table 8 shows the random part of the mixed effect models. The random effects vary across items and persons. The variance in person intercepts ($\tau_{00person}$) refers to ability, and the variance in item intercepts (τ_{00item}) is the variance in difficulty that is not explained by item characteristics. Importantly, the $\tau_{00\ person}$ and $\tau_{00,item}$ do not substantively change in the process measures models. This indicates that the process measures do not explain the variance in personal ability or item difficulty. The top of Table 8 shows the fixed part of the mixed effect models. The fixed effects are constant across persons and items.

III.4.2.6 Process measures and item characteristics.

The effect of item characteristics (topic and task type) is influenced by the process measures time-on-task and text-graph transitions. Some of the change in the likelihood of giving a correct answer can be attributed to differences in time and integration requirements for different topics or task types. When adding text-graph transitions to the model, the estimated effect of task type is lower, because the two task types have different integration requirements.

III.4.3 Discussion 1

We investigated the association of time and text-graph transitions with task success during initial reading and task completion. We hypothesized that process measures have different effects during the two comprehension phases because initial reading requires more controlled processing than task completion and task completion more automatic processing than initial reading. Therefore, during initial reading, taking more time indicates more elaborate mental model construction, and more text-graph transitions indicate global coherence formation. During task completion, on the other hand, taking a shorter amount of time indicates fluent processing of the task material and fewer text-graph transitions indicates the ability to find relevant information quickly.

In line with our hypotheses, time and text-graph transitions can have different effects on task success during initial reading and task completion. Time spent on initial reading has a marginally significant positive effect on task success, whereas time-on-task has a significant negative effect on task success. Moreover, performing transitions between text and graph during initial reading has an insignificant positive tendency, while transitions during task completion have a significant negative effect. The negative association between time-on-task and task success is in line with our hypotheses, suggesting that fluent processing is more likely to result in an accurate answer because it reflects a higher skill level. The negative association between text-graph transitions and task success during task completion allows for a more fine-grained interpretation: fluent processing involves the ability to find relevant referential connections between text and graph with few search iterations.

The effects of the process measures during initial reading were not significantly different from zero, although the effects for initial reading and task completion were significantly different from each other. Therefore, the effects of time and text-graph transitions are at least different between comprehension phases. Finally, the hypothesis that text-graph transitions improve model fit above and beyond time-on-task was not supported.

In light of our current results, it is necessary to consider that numerous studies have demonstrated that individual characteristics, such as prior knowledge and reading and graph comprehension abilities, influence comprehension (e.g., Kintsch, 1988; Ozuru et al., 2009; Shah & Freedman, 2011). Moreover, comprehension outcomes may be influenced not only by individual characteristics, but also by the relationship between process measures and task success. Prior knowledge and reading and graph comprehension could potentially change the degree to which individuals process material in a controlled or automatic way.

In summary, the overall pattern of results supports our hypotheses. However, the effects we found need to be replicated and refined with a second sample. In additions, we want to explore how individual characteristics such as prior knowledge and reading and graph comprehension ability influence the relationship between process measures and task success.

III.5 Study 2

In the second study, we first intended to replicate the first study's findings. Therefore, we used the same materials, the same data preparation, and the same statistical analysis. Second, we further included the assessment of individual characteristics (i.e., reading and graph comprehension ability and prior knowledge) in the study design in order to assess how these individual characteristics influence text-graph comprehension processes.

III.5.1 Individual characteristics and text-graph comprehension

It is uncontroversial that prior knowledge and reading and graph comprehension abilities (Kintsch, 1988; Ozuru et al., 2009; Shah & Freedman, 2011) can affect comprehension outcomes. However, we know little about how they influence the relationship between process measures and comprehension outcomes. The following section discusses the role of prior knowledge and reading and graph comprehension skills for text-graph comprehension.

III.5.1.1 Prior knowledge and the effect of process measures.

For the purpose of this study, we define prior knowledge as individuals' pre-existing knowledge related to the text content. We expect topic-relevant knowledge to have a significant influence on comprehension because the information explicitly stated in a text is often insufficient for the construction of a coherent mental model of the material; pre-existing knowledge is often required (Kintsch, 1988). Many studies have shown that prior knowledge improves text comprehension (e.g., Ozuru et al., 2009) and comprehension in general (McNamara & Magliano, 2009). In the context of text-picture comprehension, Schnotz (2005) argued that prior knowledge is the third source of information, in addition to text and graph, that helps organize mental model construction. Consequently, prior knowledge may have a mediating effect because it organizes mental model construction or because it facilitates the selection of task-relevant information.

III.5.1.2 Reading and graph comprehension and the effect of process measures.

Text-graph comprehension may require comprehension skills unique to texts and graphs. Text comprehension requires reading skills (Kintsch, 1988). Reading skills depend on efficient component processes of reading comprehension on the word, sentence, and text level (Richter et al., 2013). Similarly, graph comprehension requires graph schemata (Pinker, 1990) and knowledge about display conventions (Lowrie et al., 2012). Graph comprehension skills depend on one's ability to understand the visual-spatial array of the graph, map the spatial relations on different levels of complexity, and map this relation to proposition statements.

In sum, comprehension abilities enable individuals to process words, sentences, or spatial relations fluently. Fluent processing should reduce the time it takes to form local coherence for the text and graph. Consequently, higher comprehension abilities make processing more automatic. Therefore, the effect of time-on-task and text-graph transitions could be less positive or more negative for high-skilled individuals compared to low-skilled individuals (Goldhammer et al., 2014).

III.5.2 Method 2

The second study was conducted with a different sample of individuals and included the assessment of various individual characteristics. Apart from these aspects, we used the same material and applied the same procedure in both studies.

III.5.2.1 Sample

72 students (23 female; *M* = 20.46 years, *SD* = 3.36) from a university in southern Germany participated voluntarily in the study for course credit. Participants were psychology or cognitive science students. Data from 24 participants had to be excluded from the data analysis (the data recording for seven was interrupted, five calibrations did not work even after the third trial, twelve due to poor data quality). The same preparation criteria were used as in Study 1. Participants were invited to the eye-tracking laboratory in groups of up to 15 individuals.

III.5.2.2 Material

The materials were identical to the materials from Study 1. However, we added the following instruments to assess individual characteristics.

Prior knowledge test. The prior knowledge test was author-constructed and was designed to assess prior understanding of the three subtopics. The population dynamics test had four single-choice and two multiple-choice questions (max score of 10), while action potential and sleep cycles tests consisted of 7 single-choice items each. A total of 24 points were possible.

Text comprehension ability test. We measured individuals' reading comprehension using an analog version of the German reading speed and comprehension test (LGVT 6-12; Schneider, Schlagmüller, & Ennemoser, 2007). In this test, students were asked to read a text containing 25 gaps and decide which of three word options should fill each gap. Reading comprehension was determined on the basis of the number of correctly identified filler words.

Graph comprehension ability test. The author-designed graph comprehension test consisted of 12 bar and 12 line graphs. The graphs illustrated data for three factors (9 points). One of the three factors was relevant. The participants were asked to answer questions about the relationships between the displayed data points. Each item had four possible answers, and only one of the four answers was correct. The items were presented in 3 item blocks, which were presented in 3 different orders (Latin square). The test took about 25 minutes; no time limit was set. The maximum score was 24.

III.5.2.3 Procedure

Participants were invited into a laboratory that contained multiple mobile eye trackers. The study started with the calibration of the eye trackers. The calibration of the eye tracker was repeated when the divergence was greater than 1. However, if calibration was not successful after the third trial, participants were instructed to move on anyway. Calibration was repeated before the graph comprehension test and the text-graph integration test.

The procedure for Study 2 deviates from that of Study 1 with regard to the fact that reading comprehension, graph comprehension, and prior knowledge were assessed in addition to text-graph integration. The order of the reading comprehension, graph comprehension, and text-graph integration tests was permutated across individuals. The session ended with a short demographic questionnaire.

III.5.2.4 Process measures

The equipment and areas of interests were the same as in Study 1. The procedures for tracking rate, treating missing values, and trimming process measures were the same as in Study 1. A detailed report on the data preparation can be found in the appendix.

III.5.2.5 Statistical analyses

The analysis for Study 2 was identical to Study 1; however, we additionally tested whether the relationship between process measures and task success was influenced by individual characteristics.

We used mean-centered variables for this purpose (i.e. the mean of each variable is zero, but the unit of measurement remains minutes, number of transitions, and test scores for time-on-task, text-graph transitions, and test results, respectively).

We used the likelihood ratio test for nested models to determine whether the study data is more plausible under the assumption of interaction between individual characteristics and process measures. A likelihood ratio test compares the goodness of fit of two statistical models, one of which is the null model and the other is a particular case of the null model, called the alternative model. The test is based on a likelihood ratio, which expresses how many times more likely the data is under the assumption of one model than the other. This likelihood ratio is used to compute a p-value. We used the process measure model from the previous analysis as the null model and the model including the interaction term as the alternative model.

We apply the same procedure for all combinations of process measures and individual characteristics separately. We start by comparing the process measure model (null model; time-on-task: ToT; text-graph transitions: TGT) to a model that includes individual characteristics (alternative model) only as the main effect. The next comparison includes this as the null model and alternative models that include the interaction terms between individual characteristics and process measures during task completion, with process measures during initial reading, and both interactions. We report the most complex alternative model.

III.5.2.6 Estimates from multiply imputed data.

The multiple imputation resulted in five imputed data sets. We performed the model comparisons for each of the five imputed datasets, and combined the chi-square statistics from all five multiply imputed datasets to test our statistical inferences.

III.5.2.7 Statistical software.

The combination of chi-square statistics from the multiply imputed datasets was performed with the function ‘micombine.chisquare’ from the package ‘miceadds’ (Robitzsch et al., 2017).

III.5.3 Results 2

III.5.3.1 Descriptive.

The pattern of results was very similar to the descriptive results from the first experiment. Therefore, we only report results that deviated from the first experiment.

The overall initial reading time (105.80 seconds) was consistently higher in the second experiment (96.32 seconds). However, the number of text-graph transitions was very similar (Experiment 1: 10.04 TGT; Experiment 2: 10.89 TGT). We can only speculate that this was due to differences in perceived time pressure as a result of the different group sizes.

Overall, Studies 1 and 2 were very similar in accuracy rate, time-on-task and number of transitions; however, items related to the sleep cycle topic were solved more often on average in Study 2, while items concerning population dynamics and action potential were solved less often. Again, we can only speculate that these differences might have been caused by the activation of prior knowledge during the prior knowledge test.

III.5.3.2 Item and person effects

We analyzed the effect of item and person on the probability of correct answers. The overall accuracy rate of 51.32% corresponds to a ‘grand’ intercept estimation of $\beta_{intercept} = 0.05$, $z = 0.22$,

p = .817. However, the probability of a correct answer ranged from 13.67% to 85.66% for the person with the lowest ability working on the most difficult item, and the person with the highest ability working on the easiest item, respectively.

III.5.3.3 Process measures and the probability of correct answers.

Time-on-task. Time-on-task did not have a significant effect on task success during initial reading ($\beta_{ToT:IR}$ = 0.04, SE = 0.18, z = 0.20, p = .840). Time-on-task had a negative effect during task completion ($\beta_{ToT:TC}$ = -0.38, SE = 0.15, z = -2.46, p = .014).

The estimate of the time-on-task effect during task completion means that the probability of a correct responses decreases (from the intercept and an average person) by 41.88% when a given item is worked on for one minute longer. The negative effect of time-on-task during task completion is in line with the hypothesis that the quick identification of task-relevant information leads to correct responses. The effect on time-on-task found in Study 2 does not support the tendency found in the first study.

Text-graph transitions. Text-graph transitions during both initial reading and task completion had a significant effect on the probability of answering an item correctly. In line with our hypothesis, text-graph transitions had a positive effect during initial reading ($\beta_{TGT:IR}$ = 0.03, SE = .01, z = 2.44, p = .015), but a negative effect during task completion ($\beta_{TGT:TC}$ = -0.03, SE = .01, z = -2.73, p = .007). These results fully support our hypotheses that text-graph transitions are associated with integration processes during initial reading and disorientation and a search for referential connections during task completion.

III.5.3.4 Incremental model fit.

In the second study, text-graph transitions did add predictive value to the model. The combination of chi-square statistics for the five datasets of multiply imputed data is χ (2) = 5.54, p =.004. Text-graph transitions are informative above and beyond time-on-task, in line with our fourth hypothesis.

In the model including both time-on-task and text-graph transitions, only transitions during initial reading still had a significant effect. This means that transitions during initial reading have a significant effect ($\beta_{TGT:IR}$ = 0.04, SE = .01, z = 2.78, p = .006) on the probability of answering item correctly even when controlling for the effect of initial reading time, text-graph transitions during task completion, and time-on-task.

The results for $\beta_{TGT:TC}$ in a model with time-on-task as control were unstable due to the imputation procedure, the high correlation with time-on-task, and the relatively low correlation with other variables. The reported p-value is the upper boundary.

Table 9. Study 1. Generalized Mixed Effect Regression with Person and Item as Random Effects, and Time-on-Task (ToT), Text-Graph Transitions (TGT) during Initial Reading (IR) and Task Completion (TC) as Fixed Effects.

		Baseline				**ToT**				**TGT**				**Combined model**			
		est	SE	*z*	*p*	est	SE	*z*	*p*	est	SE	*z*	*p*	est	SE	*z*	*p*
Fixed Effects																	
ToT	IR					0.04	.18	0.20	.840					-0.31	.22	-1.42	.156
	TC					-0.38	.15	-2.46	.014					-0.04	.24	-0.18	.861
TGT	IR									0.03	.01	2.44	.015	0.04	.01	2.78	.006
	TC									-0.03	.01	-2.73	.007	-0.02	.02	-1.55	.124[a]
Item char.	Intercept	0.67	.38	1.77	.077	0.96	.50	1.94	.053	0.49	.38	1.28	.202	0.93	.50	1.87	.062
	Task type	-0.78	.34	-2.27	.024	-0.82	.33	-2.49	.013	-0.58	.33	-1.75	.081	-0.61	.36	-1.58	.115
	Action pot.	-0.43	.38	-1.12	.265	-0.37	.37	-0.99	.324	-0.21	.37	-0.57	.568	-0.16	.37	-0.44	.658
	Sleep cy.	0.13	.39	0.34	.738	0.14	.38	0.38	.707	0.19	.37	0.53	.599	0.21	.37	0.64	.526
Random Effects																	
	$\tau_{00,person}$	0.27				0.25				0.22				0.23			
	$\tau_{00,item}$	0.20				0.18				0.16				0.17			
	N_{id}	48															
	N_{item}	12															
	Obser.	576															
	Missings					ToT:IR=5; ToT:TC: 2				TGT:IR = 21; TGT:TC = 88				…			
	Deviance	755.3				685.4				681.6				678.0			

III.5.3.5 Moderation of individual characteristics

We performed multiple model comparisons to find potential effects of prior knowledge on the relationship between process measures and task success (Appendix Table 9). Prior knowledge (β_{PK} = 0.14, *SE* = .06, *z* = 2.37, *p* = .018) and graph comprehension (β_{GC} = 0.08, *SE* = .03, *z* = 2.50, *p* = .013) had a significant effect on task success. However, the effect of reading comprehension (β_{RC} = 0.05, *SE* = .04, *z* = 1.42, *p* = .158) was not significant.

The model comparisons revealed that the model including interactions between prior knowledge and time during initial reading and between prior knowledge and time during task completion had significantly better model fit. Graph comprehension did additionally explain task success in the time-on-task model, but prior knowledge and graph comprehension do additionally explain task success in the text-graph transition model.

We will report the interaction effects of prior knowledge and time-on-task for task success. Both time-on-task during task completion ($\beta_{TOT:TC}$ = 0.11, *SE* = .05, *z* = 2.14, *p* = .032) and prior knowledge (β_{PK} = -0.28, *SE* = .12, *z* = -2.341, *p* = .019) have significant main effects in the model. In addition, the interactions between prior knowledge and time-on-task during initial reading ($\beta_{TOT:IR \times PK}$ = 0.29, *SE* = .09, *z* = 3.12, *p* = .002) and task completion ($\beta_{TOT:TC \times PK}$ = -0.29, *SE* = .08, *z* = -3.51, *p* < .001) are both significant. We graph the overall moderation effect on prior knowledge in Figure 10. Extreme values for prior knowledge have been selected for instructional reasons.

Figure 10 (left) shows the estimated initial reading time effect for individuals with high and very low prior knowledge. High and low prior knowledge individuals have a similar probability of success for initial reading times below and at the average. However, when the initial reading time is higher than average, the probability of success is different for high and low prior knowledge individuals, with the probability of success increasing for individuals with high prior knowledge individuals and decreasing for individuals with low prior knowledge.

Figure 6 (right) shows the estimated task completion time effects for very high and very low prior knowledge. When task completion takes less time than average, the probability of success is high for individuals with high prior knowledge and low for individuals with low prior knowledge. In contrast, when task completion takes longer than average, the probability of success is low for high prior knowledge individuals and high for low prior knowledge individuals. At average times, both high and low prior knowledge individuals have a similar probability of success.

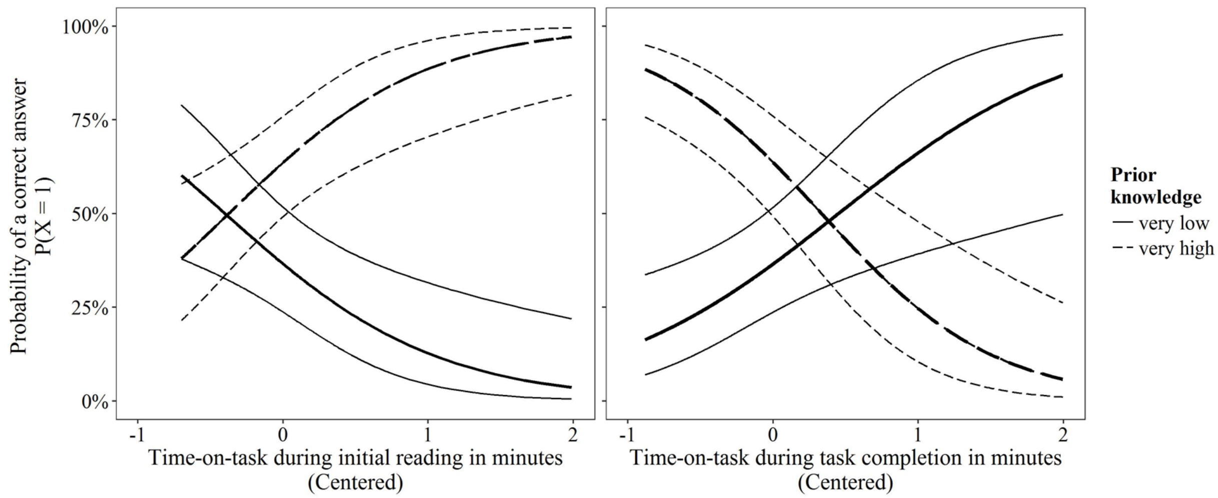


Figure 10. The plots show the estimated effect of initial reading time (left) and task completion time (right) for a person with very low (solid line) and very high (dashed line) prior knowledge. The thin lines represent the 95% confidence interval. Plots range from -1 to 2 minutes on average because the time-on-task distribution is skewed to the right.

III.5.4 Discussion 2

The second study was intended to replicate the results of Study 1. We investigated the associations of time-on-task and text-graph transitions during initial reading and task completion with task success. The findings of Study 2 mostly overlap with those of Study 1. Possibly due to the greater statistical power, we additionally found a significant effect of text-graph transitions during initial reading. This positive effect of text-graph transitions during initial reading was in line with our hypothesis and previous findings that text-graphics transitions are positively associated with learning outcomes (Mason et al., 2015; 2013). Furthermore, text-graph transitions increased model fit above and beyond time-on-task, indicating that global coherence formation and the inability to make referential connections are critical cognitive processes in addition to elaboration and fluency.

The effect of time-on-task during initial reading was again not significant. We can only speculate that this effect was not present because participants, in general, invested enough time in reading the task. Given this baseline, engagement in global coherence processing seems to instead be more crucial.

Surprisingly, even though we found no mean effect of time during initial reading, we found that prior knowledge moderated the effect of time-on-task on task success during both initial reading and task completion. Similar to Schwonke et al. (2009), we found a positive moderation effect of prior knowledge for initial reading, meaning that the relationship between visual activity and comprehension outcomes becomes more positive when prior knowledge is high.

III.6 Summary and conclusions

In the current study, we aimed to disentangle the ambiguous interpretation of process measures in text-graph comprehension by analyzing the associations between time-on-task, text-graph transitions, and task success. We made a distinction between the two comprehension phases: initial reading and task completion. We argued that the association between process measures and task success depends on the degree to which processing is automatic or controlled. The initial reading phase is assumed to be more controlled because it serves to facilitate mental model construction. The task completion phase is more automatic than initial reading because it requires the fluent selection of task-relevant information. For the initial reading phase, we hypothesized that taking more time and more text-graph transitions would lead to greater task success because they are indicative of more thorough mental model construction, i.e., elaboration processes and integration, respectively. For the task completion phase, we hypothesized that taking more time and more text-graph transitions would be indicative of a lack of fluency, i.e., the inability to find task-relevant information and a greater need to search for referential connections, respectively. Since individual differences influence processing of materials, prior knowledge and reading and graph comprehension could act as moderators.

Our results show that text-graph transitions can be positively and negatively related to task success depending on the comprehension phase. Indeed, performing many text-graph transitions can indicate either integration on the one hand or disorientation on the other hand. Text-graph transitions during the initial reading phase indicate integration and the construction of an initial mental model, but they indicate disorientation when a specific task has to be solved.

The results for time-on-task are a bit more complicated since time-on-task depends on the level of prior knowledge. Specifically, the effect of time becomes more positive with higher prior knowledge during initial reading and more positive with lower prior knowledge during task completion. We explain these results by referencing two different *mechanisms of action* of prior knowledge. On the one hand, prior knowledge may help with organization (Schnotz, 2005); on the other hand, it may help people focus on task-relevant information (Canham & Hegarty, 2010). More specifically, the construction of mental models from text-graph material requires controlled processing, but only organized controlled processing is associated with task success. Since prior knowledge helps organize mental model construction, time has a more positive effect on task success when prior knowledge is higher.

In contrast, the effect of prior knowledge is different when a task is being answered. High prior knowledge enables a person identify task-relevant information in little time. When prior knowledge is low, investing more time into the search for relevant information may result in better task success. Low prior knowledge individuals have to process the material in a more controlled way than high prior knowledge individuals to be successful. Since low prior knowledge individuals need to invest more time to be successful, time-on-task has a more positive effect when prior knowledge is lower.

We may conclude that spending time on initial reading indicates elaboration only when prior knowledge is high. Conversely, time-on-task during task completion likely reflects a person's skill level and more fluent processing; in addition, prior knowledge accelerates processing.

These results suggest that process measures can indicate different cognitive processes depending on the comprehension phase. Consequently, it is essential to conceptually separate process measures (time and text-graph transitions) and cognitive processes (i.e. elaboration, fluency, integration, and disorientation). Time-on-task cannot be equated with processing speed (Van Der Linden, 2009), and performing text-graph transitions cannot be equated with integrative processes or global coherence formation.

Separating process measures and cognitive processes may also change our understanding of the *eye-mind assumption* (Carpenter & Just, 1975). The literature has consistently demonstrated that there is a relationship between eye movements and learning (e.g., Scheiter, & Eitel, 2015) and comprehension outcomes (e.g. Schnotz, et al., 2014). This supports the *eye-mind assumption*, i.e., the idea that gazes represent engagement of the mind. However, the presence of negative and positive relationships with task success indicates that the link between gazes and *what* the mind is engaged with is not *direct*. This may be particularly true for cognitive processes on a conceptual level, like text-graph comprehension, rather than a perceptual level. Consequently, interpreting gaze behaviors requires a cognitive task analysis that considers different processing phases, individual characteristics, and item characteristics.

III.6.1 Limitations and future studies

Even though we aimed for replication, we found mean differences between Studies 1 and 2 with respect to initial reading, as well as differences in the relative difficulty of the topics (i.e. sleep cycles appeared to be easy in the second study). These discrepancies may be the result of the prior knowledge test, which was only part of study 2. The prior knowledge test may have activated prior knowledge before participants read the text and exposed participants to task-relevant terms early on. However, the results of the two studies concerning fixed effects are largely consistent, making us confident that the two studies reflect very similar cognitive processes.

There is a possibility that the process measures may have had different effects on task success depending on task type. However, we define text-graph integration as an activity that involves mappings both from text to graph and from graph to text. The two task types represent these two cognitive activities. The fixed effects we report are constant across all items and individuals. Therefore, the fixed effect estimates represent two crucial cognitive activities related to text-graph integration, which we consider more meaningful than investigating the effect of process measures for a single activity. Nonetheless, our primary item selection criteria were difficult, because this study design relies heavily on obtaining a balanced proportion of correct and incorrect answers for statistical power.

From a theoretical perspective, the non-linear effect of process measures (Naumann, & Goldhammer, 2017) on task success is highly interesting. For instance, the time-on-task effect could be u-shaped. This means spending a great deal of time on task completion may increase the probability of task success. We argue that elaboration and fluency processes can be superimposed during the task completion process, resulting in a u-shaped relationship between time or text-graph transitions and task success. In our study, we addressed this issue by separating these two processes as much as possible by introducing two comprehension phases into the design and using a sample with few individual differences and items that were similar in difficulty. However, future studies could address the non-linear effect by using items that range more in difficulty and a more diverse sample of individuals.

Chapter IV. Contrasting a text-centered versus a multiple-representations perspective

Students must mentally integrate information from both texts and graphics to comprehend illustrated science texts. Mental integration can be especially challenging when text is combined with graphs (i.e., two-dimensional displays of relationships among quantitative variables). Recent research suggests that the comprehension of text and graphs involves five sub-processes: (a) understanding the visual-spatial array of the graphs, (b) interpreting the graph, (c) comprehending relevant text passages, (d) mapping relevant text passages onto relevant graph elements, and (e) mapping relevant graph elements onto relevant text passages. Although the relevance of these sub-processes is uncontroversial, their hierarchical dependency has not yet been studied in detail. The present study investigated the dependencies among these sub-processes by contrasting a text-centered and a multiple-representations perspective. Knowledge Space Theory was used to define two different knowledge structures reflecting dependencies among the sub-processes above as postulated by the two perspectives. Fifty individuals were asked to work on five different task types to assess their ability to perform text-graph comprehension sub-processes. Results showed that the knowledge structure developed from a text-centered perspective better fitted the observed response pattern. Accordingly, text-graphics comprehension may not necessarily require comprehension of graphic and text separately; instead, text comprehension seems to serve as a prerequisite, whereas graphic comprehension may result from integrated text-graphic comprehension.

IV.1 Introduction

Combined text and graphics layouts are a major design feature in, for instance, science textbooks (e.g., Schnotz, 2014), where they serve as an effective tool to enhance learning outcomes (e.g., Mayer, 2005). However, students will benefit from illustrated texts only when they can integrate information from both text and graphics (Seufert, 2003). Recent empirical evidence demonstrates that such integration is essential for comprehending illustrated texts derived from online (i.e., eye movement data) and offline (i.e., cross-modal memory intrusions) indicators (online e.g., Mason, Pluchino, Tornatora, & Ariasi, 2013; offline e.g., Schüler, Arndt, & Scheiter, 2015). For integration, students must identify relevant information from each representation, organize relevant information into coherent modality-specific mental models for text and graphics (i.e., local coherence formation, Seufert, 2003) and then identify correspondences between text and graphics (i.e., global coherence formation, Seufert, 2003). Thus, several sub-processes contribute to the construction of an integrated mental model. The relevance of these processes is by-and-large uncontroversial; however, their dependency has been minimally investigated.

Therefore, the present article addresses this gap by evaluating two different theoretical perspectives surrounding these dependency, namely, a text-centered and a multiple-representations perspective, which differ in prerequisite relationships that they come with. These prerequisite relationships are investigated by evaluating two different knowledge structures (Falmagne, Koppen, Villano, Doignon, & Johannesen, 1990) that may underlay text-graphics comprehension by their fit with the observed response patterns of student participants on different text and graphics comprehension tasks. Hence, the goal of this study was to identify the most applicable knowledge structure for text-graphics comprehension.

We will refer to a specific type of graphics, namely, graphs. Graphs are quantitative axis diagrams that use an apposed-position language to represent relationships between data points (Lowrie, Diezmann, & Logan, 2012). Graphs use two-dimensional space to visualize relationships and to convey meaning, like schematic diagrams, line drawings and pictures of real-world objects. Hence, they can be considered depictive or analogous representations (Schnotz, 2014), in which the quantitative structure conveyed by the graph reflects the quantitative structure of what is being represented. In contrast to picture, for instance, graphs usually do not depict concrete objects; instead, they focus on conveying abstract information about quantitative relationships among variables. Therefore, they are sometimes referred to as logical pictures (Schnotz, 2014). Despite these differences, text-graph comprehension can be seen as a particular case of text-graphics comprehension, where integrating information presented in different symbol systems (i.e., symbolic text paired with analogous visual-spatial depictions) is essential for in-depth understanding.

Therefore, we will discuss graph comprehension against the background of the literature on text-graphics comprehension, except where specifics of the visual representation (e.g., processes that are relevant to graphs only) are concerned. Benefits of using graphs as compared to pictures or schematic diagrams, for instance, are that students can perform a higher number of tasks within the same amount of time and those tasks can be designed so that they require only one cognitive process in isolation (e.g., reading, processing of visual information). Both features are relevant for the present study.

IV.1.1 Processes involved in the comprehension of text and graphics

To comprehend illustrated texts, local processes, which refer to the respective representational format in isolation, and global processes, which integrate both representational formats are required (Mayer, 2005; Seufert, 2003). In this vein, text-graphics comprehension can be described as resulting from five (a-e) sub-processes.

On the one hand, local processes serve the understanding of each of the given representations (Ainsworth, Bibby, & Wood, 2002) or the selection and organization of information into modality-specific mental models (Mayer, 2005).

As such, successful local processing of texts (c) enables students to distinguish between surface and deep structure of sentences and texts. The deep structure of a sentence is a theoretical construct, which makes the underlying logical and semantic relations explicit and is independent of a specific sentence with specific syntax and specific words (Royer, Hastings, & Hook, 1979). Accordingly, someone who grasps the meaning of a sentence can match two sentences with the same deep structure even though the sentences may have different surface structures (Royer et al., 1979). This process requires a mixture of linguistic knowledge, word knowledge, and reading skills (Kintsch, 1988).

Similarly, successful local processing of graphics information (b) enables students to distinguish between the surface and deep structure of a given graph. The deep structure of a graph makes the conceptual and logical relationships between variables explicit, and independent from a specific array of points and visual features such as dots, lines, and areas (Pinker, 1990). Understanding a graph requires constructing a mental model of its content (Pinker, 1990). Therefore, the deep structure of the graph directly represents the content of the graph (Schnotz & Baadte, 2015).

We further distinguish between (a) understanding of the mere visual array of the graph (Pinker, 1990) and the (b) interpretation of the graph. Someone who understands the visual array of a graph should be able to match two graphs with the same deep structure even though they may show different point arrays and different visual features. Someone who can interpret a graph should also be able to match the meaning of a sentence to a referring feature of the graph. Both, understanding the visual array of a graph and its interpretation require graph schemata and knowledge about displaying conventions (Lowrie et al., 2012).

However, global processes have been argued to link both text and graph information at a conceptual level (Ainsworth, 2006), integrating modality-specific mental models (Mayer, 2005) and integrating a propositional representations of the text with mental models of the content (Schnotz, 2014). Successful global processing involves (d) the mapping of relevant text passages onto referring graph elements and (e) the mapping of relevant graphic onto the referring text passages (Seufert, 2003). Mapping text onto graph information requires matching the text's deep structure onto specific graphs information with a corresponding deep structure. Likewise, mapping relevant graph elements onto referring text passages requires matching the deep structure of a graph onto referring sentences of text.

In the present study, five different item types were developed that allowed the assessment of each of the five sub-processes in isolation. As mentioned earlier, little controversy surrounds the relevance of local and global processes for text-graphic comprehension. However, exactly how these processes depend on one another remains less clear. Two different perspectives on that have been proposed in the literature, which are described in the following passages.

IV.1.2 Hypothesized prerequisite relationships among text-graphic comprehension processes

According to a *text-centered perspective*, local coherence formation within the text precedes and is a prerequisite for understanding a corresponding graph. That is, information from the text is used to build local coherence within the graph, which then results in global coherence between text and graph information. This perspective is supported by empirical findings suggesting that processing of multimedia materials is primarily driven by information given in the text (Ozcelik, Arslan-Ari, & Cagiltay, 2010; Scheiter, & Eitel, 2015). For instance, Hegarty and Just (1993) showed that students who learn about the functioning of a pulley system from an illustrated text passage first read a text paragraph and then consulted the picture for information corresponding to the textual information. This sequential reading pattern should affect the dependency between local and global coherence formation. Furthermore, the text-centered perspective is in line with findings showing that comprehension of multimedia material is strongly related to students' text comprehension abilities (Scheiter, Schüler, Gerjets, Huk, & Hesse, 2014).

In contrast, the *multiple-representations perspective* assumes that local processes of text and graphics comprehension are both necessary before global processes of coherence formation can take place (Seufert, 2003). This perspective is prominent in theories of multimedia learning such as the Cognitive Theory of Multimedia Learning (CTML, Mayer, 2005). According to the CTML, relevant information from words and pictures is first selected and then organized into separate modality-specific (verbal and pictorial) mental models. Finally, these modality-specific models are integrated with each other into a single coherent mental representation under consideration of prior knowledge. Before integration, text and picture comprehension is assumed to occur in parallel and by-and-large independent of each other (cf. Eitel, Scheiter, Schüler, Nyström, & Holmqvist, 2013).

At present, it is hardly possible to distinguish between these two different pathways to text-graphics comprehension given the available empirical data. This is because local and global processes of text and graphics comprehension could not be disentangled unambiguously in the studies described above, which in turn, prevent testing the different perspectives against each other. Moreover, because the contradictions between the perspectives occur at a detailed level of analysis that pertains to the order of prerequisite relationships among sub-processes, a higher level of precision in formulating assumptions is necessary than what is typically present in multimedia research. To achieve this higher level of precision, we made use of an approach to competence modeling called Knowledge Space Theory (Falmagne et al., 1990), which allowed us to specify the knowledge structures underlying the two perspectives outlined earlier.

IV.1.3 Modeling knowledge structures underlying text-graphics comprehension using KST

Knowledge Space Theory is a set-theoretical framework that allows defining knowledge structures (i.e., knowledge elements and their prerequisite relationships). In the following, we will provide a brief introduction to this approach (for a more detailed description see Heller, Steiner, Hockemeyer, & Albert, 2006).

IV.1.3.1 Modeling of knowledge structures based on prerequisite relationships among items

A knowledge structure is defined as a pair (Q, K) in which Q is a non-empty set, and K is a family of subsets of Q. The set Q is called *the domain of the knowledge structure*. Q consists of elements that are referred to as *items*. The subsets of items in the family K are labeled knowledge states (Falmagne et al., 1990). Items are denoted with parentheses (i.e., (a)), and knowledge states are denoted with brackets (i.e. {a}). A knowledge state represents the subset of items an individual masters in the domain Q. Accordingly, the domain of text-graphics comprehension Q_{tgc} consists of five items reflecting above described sub-processes (a) understanding the visual-spatial array of a graph, (b) graph interpretation, (c) sentence comprehension, (d) mapping text onto graph information, and (e) mapping graph onto text information: $Q_{\text{tgc}} = \{a,b,c,d,e\}$.

Theoretically, an individual can generate one out of $2^5 = 32$ different response pattern when answering five items of domain Q_{tgc}, because each of the five items can either be solved or not solved. Hence, the knowledge structure of Q_{tgc} may contain up to 32 different knowledge states. However, only some knowledge states are plausible due to hierarchically increasing cognitive demands between items of the domain (Heller et al., 2006). The hierarchically increasing cognitive demands represent prerequisite relationships between items. Consequently, mastery of one item can be a prerequisite for mastery of another item. These prerequisite relationships between items must be derived from the *domain ontology* (Heller et al., 2006), which is the theoretical and empirical evidence pertaining to a specific domain and its possible prerequisite relationships. In our case, the two perspectives on text-graphics comprehension described above allow the derivation of two different knowledge structures based on different prerequisite relationships among the constituting sub-processes (a to e as measured with different items). These theoretically derived knowledge structures can be compared directly to observed response patterns, which in turn allows the validity of these two perspectives to be evaluated.[i]

IV.1.3.2 Knowledge structures regarding text-graphics comprehension

According to the *text-centered perspective* (K_{TC}) it is assumed that a graph is understood by making referential connections between text and graph information. As a consequence, (c) sentence comprehension and (b) graph interpretation is required for mapping text onto graph information and vice versa (c→d; b→d). Additionally, (a) the visual-spatial array of a graph is understood through (b) interpreting graph information (b→a). Furthermore, (e) mapping graph onto text information first requires (d) mapping text onto graph information (d→e). Figure 11 (left panel) shows the prerequisite relationships derived from the text-centered perspective as a Hasse diagram (Falmagne et al., 1990). The respective knowledge structure defined by the prerequisite relationships of K_{TC} for the domain Q_{tgc} has ten knowledge states (Figure 12, left panel):

$$K_{\text{TC}} = \{\{\emptyset\},\{b\},\{c\},\{a,b\},\{b,c\},\{a,b,c\},\{b,c,d\},\{a,b,c,d\},\{b,c,d,e\},\{Q_{\text{tgc}}\}\} \qquad (1)$$

In contrast, following the multiple-representations perspective (K_{MR}) global integration processes only take place after both text and graph have been understood separately. One must be able to (a) understand the visual-spatial array of graph before being able to (b) interpret the graph (a→b). To be able to both map text onto graph information and vice versa (d & e), one must be able to (b) interpret a graph and to (c) comprehend relevant sentences (i.e., b→d, b→e, c→d, c→e). Figure 11 (right panel) shows the prerequisite relationships derived from the multiple-representations perspective. The knowledge structure defined by the prerequisite relationships of K_{MR} for the domain Q_{tgc} has nine knowledge states (Figure 12, right panel):

$$K_{MR} = \{\{\emptyset\},\{a\},\{c\},\{a,b\},\{a,c\},\{a,b,c\},\{a,b,c,d\},\{a,b,c,e\},\{Q_{tgc}\}\} \quad (2)$$

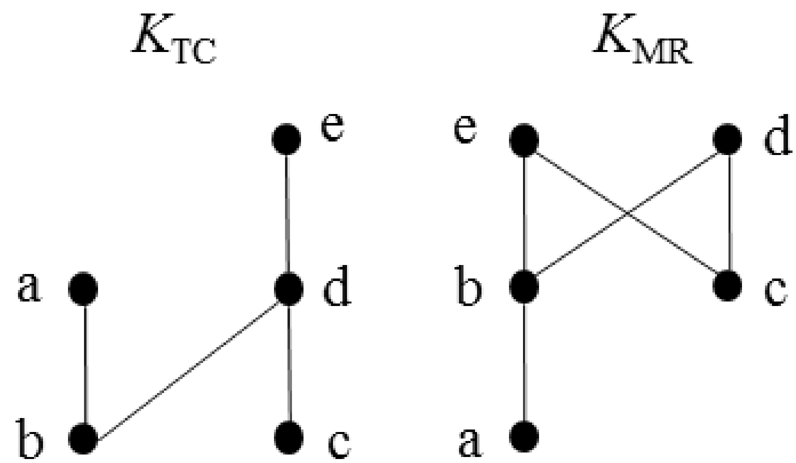


Figure 11. Hasse diagram (Falmagne et al., 1990) depicting prerequisite relationships among the five sub-processes underlying text-graph comprehension from the text-centered perspective (left panel) and the multiple-representations perspective (right panel): (a) understanding the visual array of the graph, (b) graph interpretation, (c) sentence comprehension, (d) mapping text onto the graph, and (e) mapping graph onto the text.

Knowledge structures K_{TC} (Figure 12, right panel) and K_{MR} (Figure 12, left panel) include shared and distinct knowledge states. They share knowledge state $K_{MR} \cap K_{TC}$ = {Ø}, {c}, {a,b}, {a,b,c}, {a,b,c,d}, and {Q_{tgc}}, whereas knowledge states $K_{MR} \cup K_{TC}$ = {b}, {a}, {b,c}, {a,c}, {b,c,d}, {a,b,c,e}, and {b,c,d,e} are distinct between the two knowledge structures K_{TC} and K_{MR}. The validity of the two knowledge structures can thus be tested empirically by determining whether the observed empirical response patterns to sub-processes a to e are more consistent with either one of the two knowledge structures as distinguished by the distinct knowledge states and their prerequisite relationships.

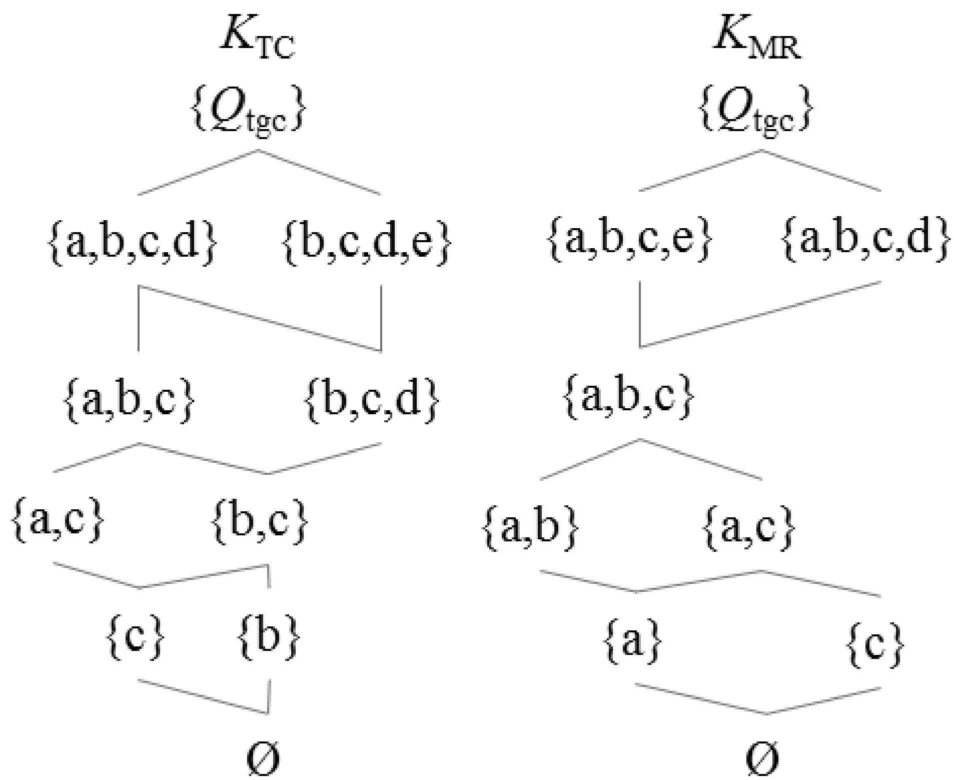


Figure 12. The different knowledge states implied by KTC for the text-centered perspective (left panel) and KMR for the multiple-representations perspective (right panel). (Ø) no sub-processes were performed, (a) understanding the visual-spatial array of the graph, (b) graph interpretation, (c) sentence comprehension, (d) mapping text onto graph information and (e) mapping graph onto text information were performed.

IV.1.3.3 Interpretation of knowledge states and response patterns

We derived knowledge structures K_{TC} *and* K_{MR}, and they included different knowledge states. To make an empirical comparison, first, we needed items capable of assessing the respective sub-processes separately. Second, observed response patterns must be assigned to a specific knowledge state. For instance, a student who masters only (c) sentence comprehension obtains the response pattern 00100. This response pattern is assigned to the knowledge state {c}. Knowledge state {c} is plausible for both knowledge structures K_{MR} and K_{TC}. On the other hand, an individual who masters (b) graph interpretation and (c) sentence comprehension obtains the response pattern

01100. This response pattern can be assigned to the knowledge state {b,c}. This knowledge state {b,c} is only included in K_{TC}. The knowledge structure that includes more of the empirically observed response patterns is considered more plausible. In fact, this means that the frequency of the response patterns 01000, 10000, 01100, 10100, 01110, 11101, and 01111 is crucial to decide which of the two knowledge structures – and thus which perspective on text-graphics comprehension – is more plausible.

However, assigning a response pattern directly to a knowledge state does not account for the possibility of response error. A response pattern may still be the result of the individual's knowledge state and response error. For instance, without response error, the observed response pattern 11010 would be associated with the knowledge state {a,b,d}. However, knowledge state {a,b,d} is not plausible for theoretical reasons, because mapping text onto a graph information requires that individuals first comprehend the referring sentence (from both perspectives). Yet, it might be that response pattern 11010 would reflect knowledge state {a,b,c,d}, but the individual missed the correct answer for (c) sentence comprehension accidentally (i.e., pattern 11110 turned into 11010). Alternatively, 11010 might reflect knowledge state {a,b} but the individual correctly guessed the answer for (d) mapping text onto graph (i.e., pattern 11000 turned into 11010).

Such influences of response error are addressed in a basic local independence model (BLIM). A BLIM constitutes a probabilistic knowledge structure and considers the effect of response error (Wickelmaier, Heller, & Anselmi, 2016). Due to this advantage, we will compare the fit of BLIMs based on both hypothesized knowledge structures in our analysis. The exact procedure will be explained in the method section.

IV.1.4 Overview of Study

In this study, we evaluated which of two the perspectives on text-graphics comprehension (i.e., text-centered vs. multi-representational) can be supported empirically. Participants completed a series of items that assessed the different sub-processes contributing to text-graphics comprehension. The comprehension test involved three different sub-domains of biology to control for effects that may only be found in a specific sub-domain. We specifically analyzed the extent that obtained response patterns matched the expected patterns for either the text-centered or the multiple-representations perspectives. The perspective that is favored by the obtained response pattern was considered more plausible.

IV.2 Method

IV.2.1 Sample

A total of 53 German adults participated in the study via Clickworkers.de. They were paid 8.50 €. Three participants did not finish the study and had to be removed from the analysis. Therefore, the analysis is based on the remaining 50 (32 males) participants (M = 34.87 years; SD = 10.58). The highest educational degrees completed by participants were lower secondary school degree (n = 8), higher secondary school degree (n = 26) and University or a University of Applied Sciences degree (n = 19). None of the university degrees were related to biology.

IV.2.2 Material

Three different topics from biology were chosen to create test items: population dynamics, action potentials, and sleep cycles. We selected these topics because they could efficiently be conveyed through both graphs and verbal explanations. Each topic addressed three core concepts. A *concept* is a central idea that is explicitly mentioned in a sentence of the text and explicitly depicted in the configuration of the graph (Appendix Table 10).

The three topics included text with 200 to 217 words. Each text had 14 sentences arranged into two paragraphs. Texts provided an overview of the interaction between predator and prey populations, the triggering of action potentials in neurons, and the sequence of sleep cycles, respectively. Examples of the respective graphs are shown in Figure 13. Based on these materials, an item pool of 81 items was created to assess participants' abilities to perform each of the sub-processes of text-graphic comprehension.

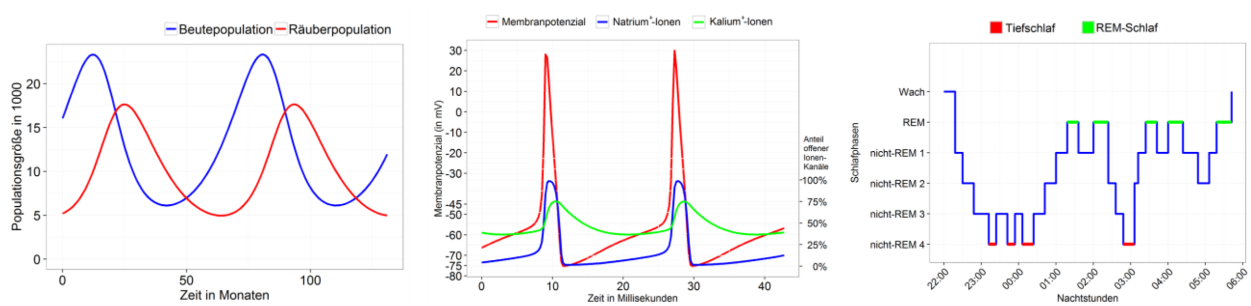


Figure 13. Examples of graphs from each topic, population dynamics (left), action potentials (center), and sleep cycles (right panel).

IV.2.2.1 Understanding of the visual-special array of a graphic (a)

Items that consisted of one reference graph and six graph options, which differed in their visual appearance, assessed participants' understanding of the visual-spatial array (See Figure 14).

Two graphs were always structurally similar to the reference graph in that they depicted identical information relevant to the meaning of the reference graphic (e.g., people enter first non-REM sleep phase after falling asleep) and differed from the reference graph only in surface features (e.g., exact length of phases), which were not relevant to its interpretation. The remaining four options not only varied in surface features from the reference graph but also in structural features (e.g., entering sleep phase). Participants were asked to identify those two graphs that showed the same structural features as the reference graph. The task was scored as correct when participants selected these two graphs. This task design did not require any semantic processing of sentence information apart from task instructions. Instead, the answer could be inferred by relying visual information of the graphs. Eighteen items were created for this task type.

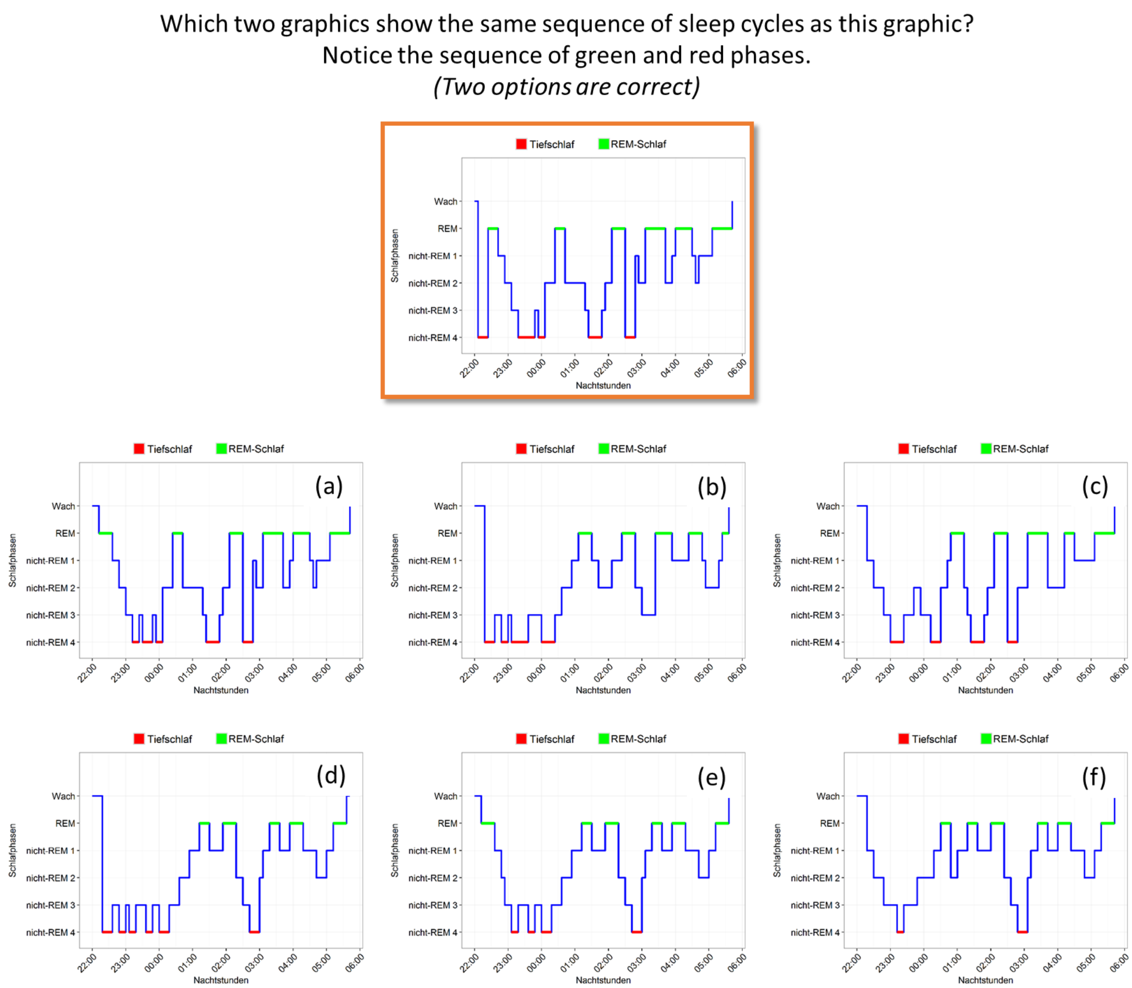
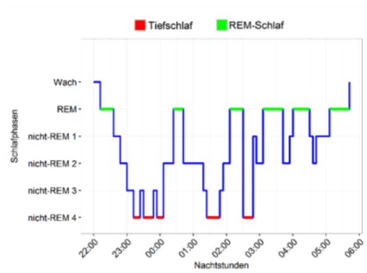


Figure 14. Sample item referring to the circadian circle on understanding the visual-array of the graphic. The top reference graphic shows that people enter deep sleep after falling asleep. Graphic options b and d show the same pattern. The entered sleep phase is the only feature that is consistent throughout at least two graphics. All other features such as the length of the sleep phases vary across all graphics. Eighteen items of this task type were created.

IV.2.2.2 Interpreting a graph (b)

Items that consisted of one reference graph and four sentence options assessed participants' abilities to interpret the content of a graph. Only one of the sentences was consistent with the content of the graph, whereas the remaining options were inconsistent. Participants were asked to select the consistent sentence. The sample item in Figure 15 shows a reference graph according to which people enter the REM-sleep phase after falling asleep. Participants selected the sentence that states that "people enter the REM-sleep phase after falling asleep". Eighteen items for this task type were created.

Which sentence describes this graphic accuratly?



After being awake people fall into to deep sleep

After being awake people fall into to REM-sleep

After being awake people fall into to frist-non-REM-sleep phase

After being awake people fall into to second-non-REM-sleep phase

Figure 15. Sample item for graph interpretation of the circadian circle. The top reference graphic shows that people enter REM sleep after falling asleep. Therefore, sentence b is correct.

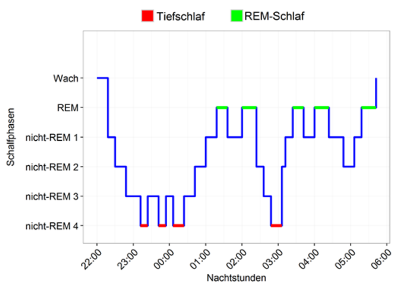
IV.2.2.3 Sentence comprehension (c)

This sub-process was assessed with a cloze test, where the missing word had to be inferred from the context of the text. In the sample item, "After falling asleep one enters the ________ sleep phase", the correct answer is the "first non-REM" sleep phase (see Figure 16.). The cloze sentence was a paraphrased version of a sentence of the corresponding text. Hence, the sentence in the text and the cloze sentence had the same structure but varied in wording and syntax structure. Eighteen items for this task type were created.

When falling asleep one enters the __________ sleep phase.

Are we sleeping deeply all night?

The activity of our brain can be measured during our sleep based on brain waves. These Brainwaves show that sleep is not a continuous state. For healthy people, different phases of sleep can be distinguished: the non-REM sleep phases and the REM sleep phases. REM stands for "Repeated Eye Movement". The REM sleep is distinguished by particularly great muscular and physiological activity and is most similar to the wake state from all sleep phases. The non-REM sleep can in turn be divided into four phases. In the first non-REM sleep phase, sleep is still very easy and becomes lower and lower over phases 2, 3 and 4.

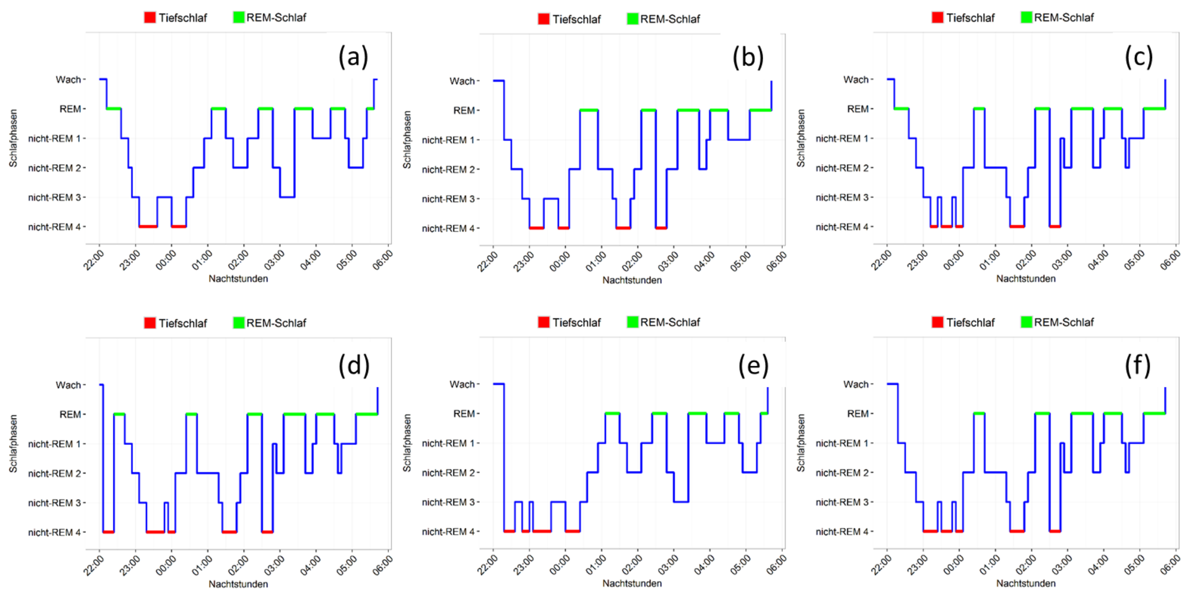


Sleep begins with the first non-REM sleep phase and then becomes ever deeper to the fourth non-REM sleep phase. The fourth non-REM sleep phase is therefore also referred to as deep sleep. The sequence from the first non-REM sleep phase to the REM sleep phase is referred to as sleep cycle. REM sleep terminates each cycle. A sleep cycle takes about one and a half hours, with a healthy person going through 4 to 6 sleep cycles in one night. The sequence of the sleeping phases and their relative share in the respective sleep cycle change during the night. The deep non-REM sleep phases (3, 4) are predominate in the first third of the night, after which the light (1,2) non-REM and REM sleep phases are predominantly. As a rule, the first two sleep cycles even include all low sleep phases of the night.

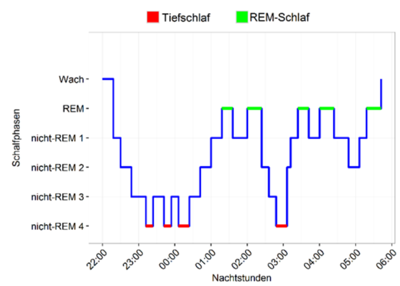
Figure 16. Depicts a sample item referring to the circadian circle. The missing word in the cloze sentence can be extracted from the material. The correct answer is "first non-REM" (all kinds of spelling were accepted, e.g., 1 non-REM, fist-non-REM)

IV.2.2.4 Mapping of text onto graph (d)

For this item type, six graphs were shown along with descriptive text. Participants had to select the two graphs that were consistent with the text. For instance, the text states that people enter the first non-REM sleep phase after falling asleep. Two graphs (a and c in Figure 17) show that the first sleep phase is the REM-phase, two graphs (b and f) indicate that the first sleep phase is a non-REM phase, and two graphs (d and e) depict that the first sleep phase is a deep sleep phase. Participants had to select graphics b and f. Nine items for this task type were created.



The activity of our brain can be measured during our sleep based on brain waves. These Brainwaves show that sleep is not a continuous state. For healthy people, different phases of sleep can be distinguished: the non-REM sleep phases and the REM sleep phases. REM stands for "Repeated Eye Movement". The REM sleep is distinguished by particularly great muscular and physiological activity and is most similar to the wake state from all sleep phases. The non-REM sleep can in turn be divided into four phases. In the first non-REM sleep phase, sleep is still very easy and becomes lower and lower over phases 2, 3 and 4.



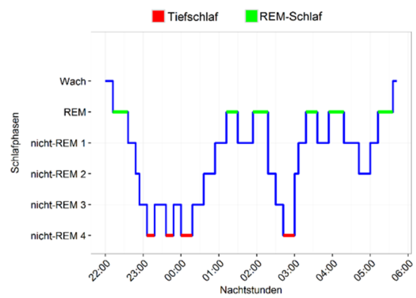
Sleep begins with the first non-REM sleep phase and then becomes ever deeper to the fourth non-REM sleep phase. The fourth non-REM sleep phase is therefore also referred to as deep sleep. The sequence from the first non-REM sleep phase to the REM sleep phase is referred to as sleep cycle. REM sleep terminates each cycle. A sleep cycle takes about one and a half hours, with a healthy person going through 4 to 6 sleep cycles in one night. The sequence of the sleeping phases and their relative share in the respective sleep cycle change during the night. The deep non-REM sleep phases (3, 4) are predominate in the first third of the night, after which the light (1,2) non-REM and REM sleep phases are predominantly. As a rule, the first two sleep cycles even include all low sleep phases of the night.

Figure 17. Sample item depicting the circadian cycle to assess the mapping of text on graphs. The six graphs showed different first sleep phases. The top left and top right images depict the REM phase as entering sleep phase. The bottom left, and bottom middle images show the deep sleep phase, and the top middle and bottom right images display a non-REM sleep phase as entering sleep phases, which would be the correct response.

IV.2.2.5 Mapping from graph information onto text (e)

This sub-process was assessed by presenting a graph that contradicted one sentence in the accompanying text. Participants answered the question by selecting the sentence in the text that was contradicted by the graph. For instance, the graph showed that the entering sleep phase is the REM-sleep phase (see Figure 8). However, the text states in one of its sentences that the entering sleep phase is the first-non-REM sleep phase. Participants had to identify this contradiction to answer the task correctly by clicking on the contradicting sentence. Eighteen items were created for this task type.

Which sentence contradicts this graphic?



The activity of our brain can be measured during our sleep based on brain waves. These Brainwaves show that sleep is not a continuous state. For healthy people, different phases of sleep can be distinguished: the non-REM sleep phases and the REM sleep phases. REM stands for "Repeated Eye Movement". The REM sleep is distinguished by particularly great muscular and physiological activity and is most similar to the wake state from all sleep phases. The non-REM sleep can in turn be divided into four phases. In the first non-REM sleep phase, sleep is still very easy and becomes lower and lower over phases 2, 3 and 4.

Sleep begins with the first non-REM sleep phase and then becomes ever deeper to the fourth non-REM sleep phase. The fourth non-REM sleep phase is therefore also referred to as deep sleep. The sequence from the first non-REM sleep phase to the REM sleep phase is referred to as sleep cycle. REM sleep terminates each cycle. A sleep cycle takes about one and a half hours, with a healthy person going through 4 to 6 sleep cycles in one night. The sequence of the sleeping phases and their relative share in the respective sleep cycle change during the night. The deep non-REM sleep phases (3, 4) are predominate in the first third of the night, after which the light (1,2) non-REM and REM sleep phases are predominantly. As a rule, the first two sleep cycles even include all low sleep phases of the night.

The graphic does not contradict the text.

Figure 18. Sample item referring to the circadian circle for assessing the mapping of graphic to text. The missing word in the cloze sentence can be extracted from the material. The correct answer is "first non-REM" (all kinds of spellings were accepted, e.g., 1 non-REM, fist-non-REM).

IV.2.2.6 Distribution of items across topics and concepts, item selection.

A total of 81 items were constructed, with 27 items referring to each topic. Nine items for each topic addressed one core concept. These always included two items for assessing the understanding of the graph's visual-array, two items on graph interpretation, two items on sentence comprehension, one item on mapping text onto graph, and two items on mapping graph onto text. Within task types, the two-item versions varied in wording or exact manipulation of the graph. For instance, regarding the entering sleep phase, the two-item versions for assessing participants' ability to map the graph information onto the text used graph that contradicted the concept by either showing the REM sleep phase or the deep sleep phase as the entering sleep phase. For mapping text onto graph information, we constructed only one variation per concept because only a limited number of graph manipulations could be performed without changing the conveyed structural features.

Wording and exact manipulation of the graphs may influence response accuracy. However, we have no prior assumption about the effect of specific wording and visual configurations. Nevertheless, they are inseparable item characteristics. To reduce the item set to five items, where one item represents one sub-process, we only considered the item version with the higher variance. We chose this item selection method because high variance items create more distinct response pattern, thus an item with higher variance is more informative than its counterpart with less variance. More distinct response patterns make it less likely to identify differences between knowledge structures.

IV.2.3 Procedure

Participants were recruited via Clickworkers.de, which is a crowdsourcing platform similar to Mechanical Turk (MTurk™). We choose Clickworker.de over MTurk™ because Clickworker.de has a larger community of native German speakers[5]. For this study, a more diverse crowdsourcing sample was particularly suited, because our data analyses rely on natural occurring differences between individuals.

Before assigning the comprehension tasks, we assessed participants' domain knowledge regarding the three topics: domain interest, academic self-concept, and their preparedness to make an effort. Participants were instructed to complete each task as accurately as possible. Prior to each new task type, a practice item was presented. Each participant only worked on two of the three topics due to time limitations.

When performing the tasks, participants first read the illustrated texts and then worked on the different items of task type a, b, c, d, and e, respectively. The order of topics as well as of items within each task type were randomized across participants. Participants were free to allocate as much time as they wanted to each task because the comprehension test was designed as a power test. The average duration of a session was M = 62.20 minutes (SD = 21.00).

---

[5] *Previous research suggested that sampling from crowdsourcing platforms can yield responses of quality comparable to those obtained from more traditional sampling methods, including data collection from undergraduate students and community samples (Follmer, Sperling, & Suen, 2017).*

IV.2.4 Statistical analyses

All analyses are based on six item sets that each address a concept within one of the two topics encountered by an individual. This leaves us with 50 (individuals) x 6 (item sets) = 300 response patterns. Analyses were performed in the R environment (R Core Team, 2012). We elaborate on model specification and model selection in the following sections. Computationally, log-likelihood and response error of each model were estimated with the 'blim' function of the R package psk (Wickelmaier et al., 2016).

IV.2.4.1 Model specification

We analyzed response pattern using a basic local independence model (BLIM). A BLIM calculates the probability of the response pattern given the knowledge state of the respective knowledge structure while considering guessing and slipping (Wickelmaier et al., 2016). The response error is defined for each item type. Response error can be fixed to a certain value or estimated. More formally speaking, we assume that responses are stochastically independent over item types q and that the response to each item type q only depends on the probabilities β_q slipping, η_q guessing and the knowledge state K of a person. The probability of the response pattern R given the knowledge state K is determined as follows (Heller & Wickelmaier, 2013):

$$P(R|K) = \prod_{q \epsilon K/R} \beta_q \prod_{q \epsilon K \cap R} 1 - \beta_q \prod_{q \epsilon K/R} \eta_q \prod_{q \epsilon Q\ (R \cup K)} 1 - \eta_q \qquad \text{F5}$$

For the following analysis, we specified the guessing parameter η_q, based on the number of response options an item type offers: (a) understanding of the visual-spatial array of a graph: η_a = 3% (two out of six), (b) for graph interpretation: η_b = 25% (one out of four), (c) sentence comprehension: $\eta_{c\,=}$ 0% (open answer), (d) mapping text onto the graph: η_d = 3% (two out of six) and (e) for mapping graph onto text: η_e = 7% (one out of 14). We did not specify a slipping parameter β_q but estimated it. The response error rates influence the maximum likelihood estimation of a BLIM, in the way that likelihood increases with greater response errors. To avoid inflation of response errors we used the minimum discrepancy maximum likelihood (MDML) estimation method. MDML uses a trade-off between the minimum discrepancy method that optimizes model parameter by minimizing the number of expected response errors, and the maximum likelihood method that optimizes model parameter by maximizing the likelihood of the model (Heller & Wickelmaier, 2013).

Additionally, we specified two reference knowledge structure denoted as K_{NULL} and K_{POWER}. K_{NULL} assumes a restrictive sequential order of the item types (i.e., a→b→c→d→e). This sequential order determines on the observed overall difficulty of item types. K_{NULL} assumes merely that less difficult item types serve as the prerequisite for more difficult item types. In case the observed response pattern are less likely for K_{MR} and K_{TC} than for K_{NULL}, response patterns can simply be explained by differences in difficulty between item types. K_{POWER} is the power set of Q_{tgc} and assumes that there are no prerequisite relationships between the items types. K_{POWER} is the least restrictive knowledge structure. When the observed response pattern are less likely for K_{MR} and K_{TC} than for K_{POWER}, response pattern can be explained without any dependency between the item types.

IV.2.4.2 Model selection

We used the Akaike Information Criteria with correction for finite sample sizes (AIC_c; Wagenmakers, & Farrell, 2004; see Appendix formula 1) as a goodness-of-fit measure. AIC_c considers a trade-off between good fit to the response pattern and the number of parameters that are necessary to fit the response pattern. The number of parameters in BLIM is related to the number of knowledge states in the knowledge structure. A BLIM with a knowledge structure that encounters all possible knowledge states would by definition fit perfectly to the response pattern. As a consequence, a knowledge structure that with more knowledge states always fits better, but may capitalize on response error. For this reason, we used the AIC_c to select the knowledge structure that fits well to the response pattern but only considered as many knowledge states as necessary. We used AIC_c instead of AIC because the number of observations divided by the number of parameters was smaller than 40 (Burnham & Anderson, 2003). AIC_c is used to calculate the fit difference between the BLIMs and the Akaike Weights (ω_i; see Appendix formula 2).

IV.3 Results

IV.3.1 Descriptive

Correct solutions were given for 51.00% of understanding the visual-spatial array items, 74.67% of the graph interpretation, 76.67% of the sentence comprehension, 71.67% mapping text onto graph, and 33.33% mapping graph onto text items. These results show that the item types address different aspects of text-graph comprehension and are overall neither too easy nor too difficult for the participants.

Figure 19 depicts the frequency of each response pattern grouped by association to knowledge structures. The majority of observed response patterns (69.00%) can be assigned to at least one of the hypothesized knowledge structures. A proportion of 39.33% of response patterns can be assigned to K_{MR}, whereas 65.33% can be assigned to the text-centered perspective K_{TC}. Accordingly, the descriptive results favor K_{TC}, because the frequency of response patterns assigned to K_{TC} is higher than the frequency of response patterns associated with K_{MR}. Also, 35.67% of response patterns could be assigned to either knowledge structure. In the following, we used the BLIM to test the empirical relevance of this difference.

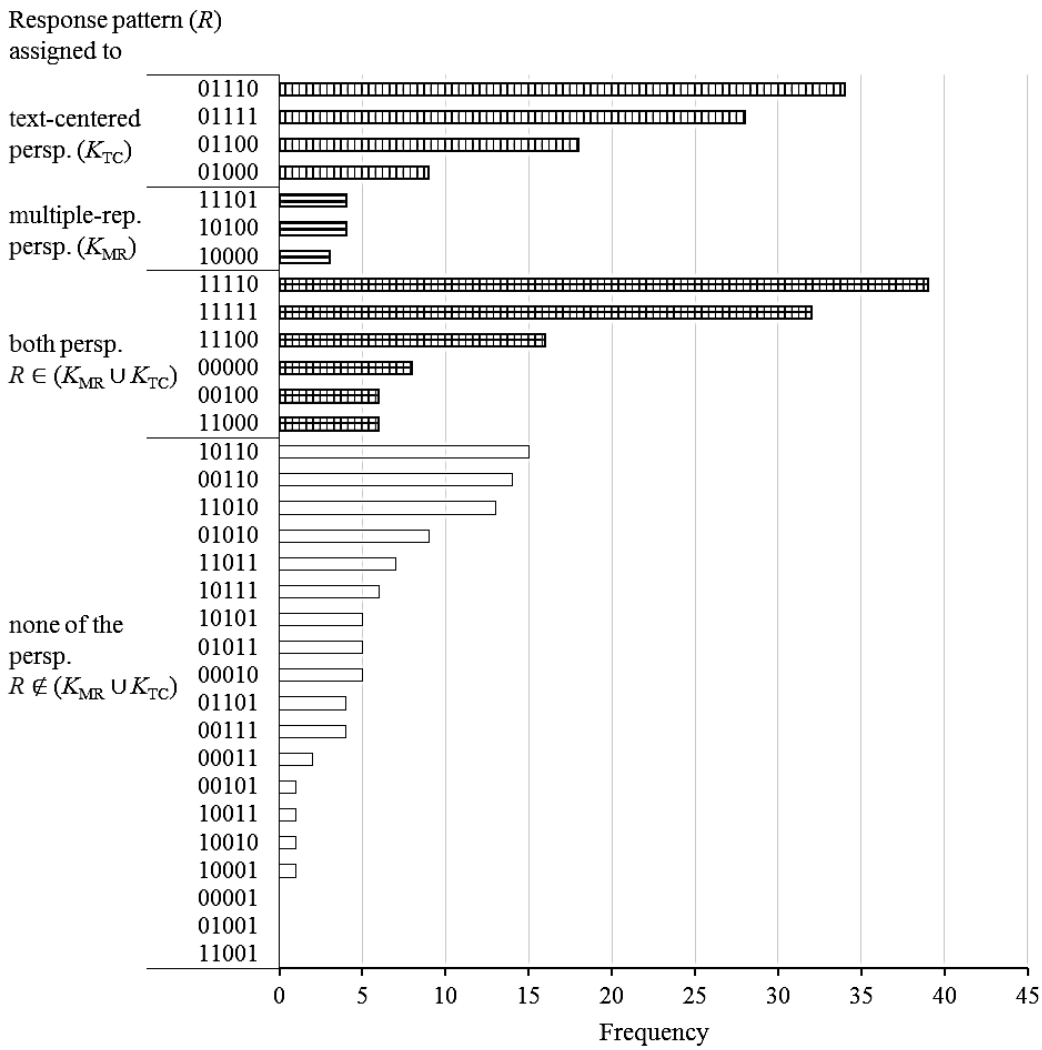


Figure 19. Frequency of response patterns R. Grouping shows response patterns specifically associated with the knowledge structure representing the text-centered perspective (TC) $R \in (KTC \backslash KMR)$, with the knowledge structure representing the multiple-representation perspective (MR) $R \in (K_{MR} \backslash K_{TC})$, with both $R \in (K_{MR} \cup K_{TC})$ or none of them $R \notin (K_{MR} \cup K_{TC})$.

Participants reported medium level of interest in biology (*M* = 57.16, *SD* = 28.79) and high preparedness to make an effort, *M* = 23.02 (*SD* = 2.18) on a scale ranging from 0 – 24. They reported relatively high academic self-concept, *M* = 11.20, *SD* = 3.08 on a scale ranging from 0 – 18. Academic self-concept, $r(48) = .083$, $p = .56$, and topic interest, $r(48) = .17$, $p = .25$, were not significantly correlated to text-graphics comprehension. However, in line with prior research on comprehension, prior knowledge (e.g., Kintsch, 1988) was significantly associated with overall text-graphics comprehension, $r(48) = .50$, $p < .001$.

IV.3.2 Basic local independency model (BLIM) selection

To determine which of the knowledge structures best fits the observed response patterns, response errors, the value of the log-likelihood function, and AIC_c of the BLIM for K_{NULL}, K_{MR}, K_{TC} and K_{POWER} were estimated first. Second, we calculated Δ_i, and ω_i, based on the BLIM's AIC_c. Table 10 gives an overview of all indicators. Response errors estimated for slipping were .49, .54, .33, 0 for K_{NULL}, K_{MR}, K_{TC}, and K_{POWER} respectively. Response errors for guessing were .10, .11, .09, and .13 for above models. K_{TC} had the largest maximized log-likelihood value ($logLike_{TC}$ = -907.63) and smallest overall response error. The log-likelihood value of K_{MR} ranked second ($logLike_{MR}$ = -923.48). Furthermore, the BLIM of K_{MR} returned the largest over all response error. The BLIM of K_{NULL} was the least likely ($logLike_{NULL}$ = -931.43). Furthermore, results for AIC_c substantiated those observed for the log-likelihood value; K_{TC} (AIC_{TC} = 1844.74) had the smallest AIC_c, followed by, AIC_{POWER} = 1863.72, AIC_{MR} = 1874.23 and AIC_{NULL} = 1883.61. In addition, all models were substantively different from one another ($\Delta_i > 10$). Consequently, ω_i values indicated decisive evidence in favor of K_{TC} relative to the other BLIMs of K_{POWER}, K_{MR} and K_{NULL}. In fact, K_{TC} received about 99% of the total weight of the considered models. However, it is important to note that ω_{TC} is not an evaluation of fit or explained variance, it only represents conditional probability.

Table 10. Model summary for BLIM of K_{NULL}, K_{MR}, K_{TC}, and K_{POWER}. [6]

BLIM	**Response error (Slipping - Guessing)**	**logLike**	**Number of Parameters (k)**	**AIC_c**	**Δ_i**	**ω_i**
K_{NULL}	.49 - .10	-931.43	10	1883.61	38.87	> .001
K_{MR}	.54 - .11	-923.48	13	1874.23	29.49	> .001
K_{TC}	.33 - .09	-907.63	14	1844.74	-	< .999
K_{POWER}	.0 - .13	-890.79	36	1863.72	18.98	> .001

Note. BLIM = basic local independency model, logLike = value of the maximized log-likelihood function, AIC_c = Akaike Information Criteria with correction for finite sample sizes, Δ_i = AIC_c difference, ω_i = Akaike weights (conditional probability of the model).

[6] *The same analysis was performed without fixed guessing rates to check for robustness of model selection. In this case, differences between models were reduced. However, model selection still strongly favored KTC. The consistent difference indicated that model differences were not artifacts of the fixed parameters.*

IV.3.3 Basic local independency model of the text-centered perspective

The BLIM for K_{TC} had an overall mean error of 41.58%. This total error could be separated into error caused by guessing (8.97%) and slipping (32.60%). The estimated slipping parameter for understanding the visual-array was β_a = 0%, for graph interpretation β_b = 18.70%, for sentence comprehension β_c = 13.92%, the mapping from text to graph β_d = 2.62%, and for mapping for graph to text β_d = 0% as well. Estimated slipping parameters for understanding the visual-array of the graph and mapping from graph onto text were very low. This suggested that the knowledge structure K_{TC} represented the response pattern best when it is assumed that individuals who understood the visual-array of the graph and were able to map the graph onto the text always answer the respective items correctly. The larger slipping parameter for graph interpretation suggests that the knowledge structure represented the response pattern best when it is assumed that some participants do not answer these items correctly, even though they could.

Based on the response patterns and the estimated response errors, the BLIM of K_{TC} estimated the probability for each knowledge state. Table 11 shows the proportion of knowledge states in the probabilistic knowledge structure K_{TC}.

Table 11. The proportion of knowledge states in the probabilistic knowledge structure K_{TC}.

Knowledge states in K_{TC}	**%**
{a,b,c,d}	20.27
{*Q*}	16.55
{b,c,d}	15.17
{b,c,d,e}	13.33
{a,b,c}	7.22
{b,c}	6.67
{b}	5.89
{Ø}	5.44
{a,b}	4.94
{c}	4.50
total	100.00

Only 5.44% of the assessed knowledge states indicate a complete lack {Ø} of all sub-processes meaning that in these cases none of the items would have been solved correctly. On the other hand, in 16.55% of the assessed knowledge states, all processes {*Q*} were performed correctly meaning that in these cases local and global processes of text-graphics comprehension were fully mastered. Knowledge state {c}, which suggests that a person is only able to understand the text but

fails in comprehending graph information and relating it to text, was the least frequent, followed by the knowledge state {a, b}, which reflects understanding of the visual-spatial array and graph interpretation. Accordingly, it hardly occurred that participants were able to perform local processes of understanding the visual array of the graph, while not understanding the text at all and being unable to relate text and graph information to each other. The most frequently assessed knowledge state was {a, b, c, d} suggesting that in most cases participants were able to perform all but one process, namely, mapping graph information onto the text. This suggests that albeit being seemingly symmetrical, mapping of text information onto a graph and mapping of graph information onto text seem to impose different challenges, a finding that will be addressed in the discussion in more detail.

IV.4 Discussion

In the present study, we investigated which sub-processes are necessary to achieve integrated text-graphics comprehension. We distinguished between five different sub-processes to capture the fine-grained structure of text-graphics comprehension: (a) understanding the visual array of the graph, (b) interpret the graph, (c) comprehending relevant text passages, (d) mapping relevant text passages onto relevant graph elements, and (e) mapping relevant graph elements onto relevant text passages. We were specifically interested in how these sub-processes depend on one another. We hypothesized their prerequisite relationships to reflect either a text-centered or a multiple-representation perspective. Using Knowledge Space Theory (Falmagne et al., 1990), we provided substantial evidence in favor of the text-centered perspective in that knowledge states specific to the text-centered perspective occurred more frequently than those specific to the multiple-representation perspective. Furthermore, the text-center perspective out performed two reference model that assumed stronger (K_{NULL}) and weaker (K_{POWER}) prerequisite relationships.

The text-centered perspective postulates that local coherence formation within the text precedes and is a prerequisite for understanding graphics. In other words, information from the text is used to build local coherence within graphics, which then results in global coherence between text and graphics. Hence, text-graphics comprehension like text comprehension alone seems to be driven by a sequential integration process (Kintsch, 1988) that involves both text and graphics information. Comprehension of relevant text passages appear to be prerequisite for mapping relevant text passages onto relevant graph elements. This comprehension then serves as a prerequisite for mapping relevant graph elements back onto relevant text passages because the integration cycle is driven by the text. Text drives the integration cycle because of its sequential nature. In line with this idea, Zwaan and Radvansky (1998) argued that graphics are “jointly incorporated with information derived from the text into an integrated situation model” (p. 164). This text-centered perspective is also supported by the sequential gaze behavior observed during comprehension of text and graphics (Hegarty & Just, 1993) as well as by the strong association between students’ text and multimedia comprehension ability (Scheiter et al., 2014). In line with the text-centered perspective, Schnotz and Wagner (2018) recently argued that text-pictures comprehension is “inherent asymmetry because text and pictures serve fundamentally different but complementary functions.” (p.1). In conclusion, the present study provides further evidence that situation models – a construct initially introduced to explain higher-level text comprehension (Zwaan & Radvansky, 1998) seem to be a parsimonious and accurate way to explain not only text but also text-graphics and thus multimedia comprehension.

IV.4.1 Implications for instruction

These insights regarding individual performance concerning the five sub-processes of text-graphics comprehension and their prerequisite relationships might be exploited to develop adaptive learning technologies (e.g., Aleven, McLaughlin, Glenn, & Koedinger, in press). In particular, personalized learning environments might adjust instructions according to individuals' level of text-graphics comprehension. The knowledge structure for graphics comprehension emphasizes (1) the importance of comprehension sub-processes and (2) their dependency. Therefore, a personalized adaptive learning environment may give instructional support that targets specific sub-processes. For instance, students may struggle to establish referential connections between graph labels and relevant text passages. As suggested by recent studies, these students may benefit from color coding that emphasizes references between text and graph information (Richter, Scheiter, & Eitel, 2016). Moreover, students who have trouble understanding the visual array of the graphic may benefit from instructions on displaying conventions (Lowrie et al., 2012). Furthermore, students who struggle to map textual onto graphical information on a conceptual level or to comprehend the relevant sentence may lack word knowledge and therefore benefit from more in depth explanations of domain-specific terms in the text (Kintsch, 1988). Finally, students who find it difficult to map graphical onto textual information on a conceptual level may benefit from explanations of terms in graph reading (Friel, Curcio, & Bright, 2001).

Additionally, the observed prerequisite relationships imply optimal learning paths (Heller et al., 2006). This means that when students struggle with more than one sub-process, instructional support should be given in a specific order. The text-centered perspective implies that instructions aiming to improve text-graphics comprehension would not be useful when text comprehension in itself is an issue. After text comprehension would be established, text-graphics comprehension can best be supported by facilitating mapping from text onto graph information, for instance, using signaling text-graphic relations (Richter et al., 2016) or by asking students to additionally draw visual elements (Schmidgall, Eitel, & Scheiter, in press).

IV.4.2 Limitations and perspective

Some response patterns associated with neither of the hypothesized knowledge structures were more frequently observed than response patterns reflecting the multiple-representation perspective. This finding seems to indicate that there may be a yet unknown knowledge structure that fits the data better than the ones we proposed and investigated in the present study. However, we are confident that the confirmatory logic underlying our deduction of knowledge structures from theory and previous research is a strength of the present study because it allows for testing explicit hypotheses. In fact, future studies are needed to evaluate whether the knowledge structures observed in the study beyond the ones we hypothesized can be replicated.

Another important point to consider in the context of text-graphics comprehension is participants' prior knowledge, which also became evident in the present study. We were not able to investigate the mechanism underlying this effect because the present study was primarily designed to contrast the two theoretical perspectives on text-graphics comprehension. These perspectives do not differ in terms of possible effects of prior knowledge. Nevertheless, Knowledge Space Theory provides a statistical framework to investigate effects of prior knowledge in more detail. Following the logic of the present study, a minimum of prior knowledge might be a prerequisite for comprehension. In future studies, we would like to focus on the effects of prior knowledge by differentiating between prerequisite knowledge (i.e., knowledge that is necessary to comprehend), domain knowledge (i.e., knowledge useful for comprehension, but not necessary), and assessed knowledge (i.e., knowledge reflected in answers to comprehension tasks).

Chapter V. General discussion

Being able to understand visualizations of data and especially graphs became an important 21st century skill (Ananiadou & Claro, 2009). Graphs represent quantities via a ‘paired with’ relation, whereas greater quantities is represented by more of some visual dimension (e.g., area, lines, diameter, angle, and color; Kosslyn, 1989). Large-scale studies found that students’ struggle to understand graphs (TIMSS, 2013). Some researchers even arguted that graphicacy is equal in status to literacy and numeracy (Åberg-Bengtsson & Ottosson, 2006). However, compared the research in reading and mathematics, relatively like is known about the underlying principals and cognitive mechanism which constitute the ability to understand graphs. This thesis investigated the underlying comprehension processes of graphicacy.

Previous research on the ability to understand graphs was channeled in two research communities: A literacy and a comprehension research community. On one hand, literacy research describes how individuals in a relevant population master realistic graphicacy tasks. On the other hand, comprehension research explains the comprehension processes in graphicacy tasks. This thesis proposed a Process-Oriented Model of Graphicacy (POMoG) to integrate perspectives from both research communities. The POMoG explains item responses in graphicacy tasks as a result of comprehension processes (e. g., mapping and visual imagery) which construct internal representations (i.e., internal representation of task, graph and content). The comprehension processes are influenced by individual characteristics (e.g., knowledge and skills), item characteristics, (e.g., complexity and graph types) and their interaction.

The POMoG can be summarized by five major assumptions. First of all it assumes, that individual differences in graphicacy are manifested in differences in comprehension processes. Comprehension processes manifest individual differences because a correct items response requires an internal representations of the task, the graph, and the content. (2) These internal representations are constructed by the comprehension processes. (3) The comprehension process consist of different process component (i.e. visual, visual-imagery and mental process). (4) The process components are influenced by individual and item characteristics and their interaction. Finally, (5) Comprehension processes are indirectly related to process measures because the interpretation of process measures depends on the relationship between the process measure and comprehension success. The studies in Chapter II, III, and IV aimed at refining the main assumption of the POMoG. Chapter II addresses the influence of Basic Numerical Abilities (BNAs) on graph reading performance, Chapter III addresses the association between time-on-task and text-graph transitions with comprehension success across comprehension phases, and Chapter IV the hierarchal dependency between comprehension processes in text-graph comprehension.

Chapter II investigated the influence of BNAs on graph reading performance. The influence of BNAs on graph reading performance explains comprehension processes because BNAs can be associated to the specific process components of the comprehension process. Therefore, the differential influence of BNAs can be used to make inferences about underlying mechanism of graphicacy performances. The influence of BNAs on graph reading performance was determined with a multiple regression analysis. In addition to BNAs, the influence of general cognitive ability, age, and gender was considered as control variables. The analyzed sample consisted of 750 German students (grade nine to eleven). The results showed that general cognitive ability was the strongest predictor of graph reading performance. More important, beyond general cognitive ability, performance in number line estimation, subtraction and conceptual knowledge about arithmetic operations (CKAO) were significant predictors of graph reading performance. The results suggested that the influence of number line estimation, subtraction and CKAO can be attributed to different underlying mechanisms. Subtraction facilitates graph reading performance because it aid the performance of arithmetic calculations and number line estimation because it aids comparisons and proportional judgments. Both subtraction and number line estimation facilitate specific process components. In contrast, CKAO may function through the control of process components. CKAO enables students to use problem solving strategies which are more effective or more efficient. For instance, replacing complex calculations by proportional judgments, as examined by Gillan (1995). Chapter II showed that BNAs are relevant individual characteristics which determine graphicacy performance even in secondary education. Improving students BNAs can potentially aid graphicacy performance. Furthermore, improving individual characteristics like CKAO which function though the control of process components may be even more beneficial.

Chapter III investigated the relationship between process measures and comprehension success across different comprehension phases. The comprehension phases are either initial reading in which processing is coherence-oriented or task completion in which processing is task-selective. The studies of Chapter III considered time-on-task and eye-movements as process measures. In multimedia research it became apparent that transitions between text and graph can be interpreted in two opposing ways. Text-graph transitions can be interpreted as integration of information from text and graph or as disorientation, the inability to find relevant information. However, the association between the text-graph transitions and comprehension success indicates whether individuals integrated information or were disoriented. Additionally, it was hypnotized that the comprehension phase influences relationship between process measures and comprehension success because they required either coherence-oriented or task selective processing. The relationship is assumed to be positive in the initial reading phase because processing serves coherent mental model construction. On the contrary, the relationship is assumed to be negative in the task completion phase because processing serves selection of task-relevant information. In two studies time-on-task and text-graph transitions from in total 77 university students that worked on twelve text-graph integration items were analyzed. The analysis was conducted with the EMPPI. The results of two studies demonstrate that spending more time and performing more text-graph transitions, can be positively associated to comprehension success during initial reading while being negatively associated with comprehension success during task completion. Moreover, content knowledge moderated the effect of time during initial reading and task completion. The results indicate that the interpretation of process measures is relative to comprehension phase and to the extent of content knowledge. Furthermore, the double sided mediating effect of content knowledge on time-on-task during initial reading and task completion indicates that content knowledge either facilitates comprehension by structuring and controlling mental model construction or by aiding the search for task relevant information.

Chapter IV investigated the hierarchical dependency of comprehension process in comprehension of text and graph. POMoG stats that responses can only be correct when individuals have an IR of the task, the graph and the content. However, it is unclear how comprehension process depend on each other when they construct internal representations. In multimedia research, there are two opposing perspectives about the hierarchical dependency of comprehension processes. From a text-centered perspective, IR of the graph is constructed as a result of an integrated comprehension of text and graph. From a multiple-representation perspective, the IR of the graph is a prerequisite for an integrated comprehension of text and graph. For this study response pattern of 50 adults that answered a large number of text-graph integration items was analyzed. Knowledge Space Theory was used to define two different knowledge structures reflecting the prerequisite relationships among the comprehension processes as postulated by the text-centered and the multiple-representation perspective. Results showed that the text-centered perspective better fitted to the observed response pattern. Accordingly, text-graph comprehension may not necessarily require comprehension of graph and text separately; instead, text comprehension seems to serve as a prerequisite, whereas graph comprehension may result from integrated text-graph comprehension.

The implications for the POMoG are discussed in the following sections, after a brief examination of strength and limitations of the thesis. Furthermore, the methodological implications and practical implications are exhibited.

V.1 Strength

The strength of this thesis is the interdisciplinary integration of research on a theoretical and a methodological level. Additionally, the modeling approaches are carefully selected based on substance and methodological aspects and the studies maximize the benefit for theory development because they rigorously deduce alternative hypothesis from contradicting theoretical perspective.

The first strength of this thesis is the integration of research from different disciplines, embodied in the Process-Oriented Model of Graphicacy (POMoG). The POMoG combines the strength from the literacy and comprehension community. The literacy research describes 'real-world' graphicacy, while experimental comprehension research explains underlying cognitive mechanism of graph comprehension, ultimately, the POMoG explains the underlying cognitive mechanism of real-world graphicacy. Since, the POMoG combines the strength of both communities, the POMoG can function as a translation tool for researchers across communities. Moreover it creates a common terminology (Section I.1), integrates the model of comprehension of visual displays (Section I.3.1) and the model of human interaction with graphs (Section I.3.2), and considers the interaction between individual and item characteristic. The POMoG offers explanation for how item difficulty is manifested in comprehension processes, and how inferences about comprehension process can be made based on process measures and comprehension success. Meanwhile, the POMoG is measurable because it formulates the data sources and levels of analysis for all of its components (i.e. comprehension processes, internal representation individual and task characteristics). Finally, the integration of descriptive and explanatory aspects can surf as a template for other research objects which are investigated from a primary 'descriptive' differential perspective and an 'explanatory' experimental community.

The second strength of the thesis is the selection and development of well-suited statistical modeling approaches. The POMoG considers individual characteristics assessed by test and questionnaires, task characteristics assessed by cognitive task analysis, and process measures, for instance, response times and eye-movements as influential factors on graph comprehension. The subsequent challenge is to jointly model these data sources. On one hand the EMPPI and on the other hand KST are modeling approaches that enable inference about comprehension processes and internal representations based on the different data sources. In detail, the EMPPI was developed in conjunction with the POMoG and combines process measures with modeling approaches from item response theory. Therefore, this thesis does not only integrate literacy and comprehension research of a theoretical level with the POMoG, but integrates research communities on a methodological level with the EMPPI. Moreover, the POMoG addresses the issue of hierarchical dependency between internal representations. The KST was selected because it is the only modeling approach that considers hierarchical dependencies. KST is a promising and underappreciated modeling approach, since hierarchical dependencies and prerequisite relationships are frequently discussed topics in educational research. Therefore, this thesis developed new and spots potent modeling approaches that are currently only discussed in methodological journals and evaluate their potential for educational psychology.

A third strength of the thesis is the rigorous deduction of alternative hypotheses. Research is most beneficial for theory development when two theories came to different hypotheses. The thesis presents two studies in which two alternative hypothesis compete in an empirical studies. In the Chapter III, process measures as indicator of integration or disorientation and in the Chapter IV, response pattern are explained as a result of a text-centered or a multiple representation perspective.

V.2 Limitations

The limitation of this thesis concern the studies samples which do not integrate representative samples with fine-grain process data collection, the absence of a measure for 'pure' graphicacy and the lag of replication studies.

The POMoG attempted to combine the strength of graphicacy and graph comprehension research. The strength of graphicacy are representative sample and realistic task and the strength of graph comprehension the collection of process data and controlled stimuli that allow inference about underlying comprehension processes. Strictly speaking, combining both means to collect process data on realistic tasks in a representative sample. However, the studies presented either analysis response from representative samples or analysis of process data from homogenous samples. In this regard, the thesis do not advance previous research. In the end, there is not just a theoretical and a methodological barrier, but also a practical barrier which separates research communities. The practical barriers is a result of resource allocation. Resources are either spend to collect cost and time efficient test data in large samples or spend on relatively expensive technical devices like eye-trackers, laboratory space and skilled examiners. The thesis did not overcome this barrier due to resource limitations.

Furthermore, the thesis emphasized the importance of putting more effort in teaching graphicacy. The benefit of teaching graphicacy depends on the degree to which graphicacy skills generalize to science, mathematics, and reading literacy. A 'pure' measure of graphicacy would be needed to investigate to influence of graphicacy on the other literacy constructs. Assessing pure graphicacy implies that test items do not involve reading or concepts from science and mathematics. This thesis did not end-up developing a standardized measure of graphicacy. However, after examining the present literature it was concluded that graphicacy item are challenging because task demand involve reading and mathematical or science concepts. Therefore, instead of excluding reading, science and mathematics from the graphicacy tests they were intentionally included. Therefore, a limitation of the thesis is that there is no standardized test of graphicacy was created, however, the test items that were created involved other task demand to be able to study comprehension processes.

Each Chapter of the thesis focuses a different aspect of the POMoG, however, the Chapters do not immediately build on each other. The POMoG proposed multiple novel assumptions. Therefore, it was attractive to focus on different assumptions and apply the related methodological approaches. This resulted in three chapter that focus on different assumptions of the POMoG and apply different methodological approaches in each study. Even though Chapter III includes two studies which build on each other, the thesis could have benefited from repeating more studies to refine and replicate results. However, future studies have to demonstrate the resilience of results.

V.3 Theoretical implication

The presented studies examined the underlying comprehension processes which lead to individual differences in graphicacy performances. The studies make inference about comprehension processes based on the influence of BNAs, based on the association between process measures and comprehension success across comprehension phases, and based on prerequisite relationships between comprehension processes in comprehension of text and graph. The result can be used to refine the assumptions of the POMoG.

According to the POMoG, comprehension processes can be influenced by individuals characteristics based on three underlying mechanisms: Substituting, fluency and control of process components. In Chapter II and III these mechanisms can be attributed to specific individual characteristics that influence graphicacy performances. Chapter II found that the BNAs, subtraction, CKAO and number line estimation influence graph reading performance. Subtraction and number line estimation are assumed to influence the fluency of process components. More specifically, subtraction should facilitate the performance of arithmetic operations, whereas, number line estimation should facilitate comparisons between distanced points and imagery comparisons (i.e. comparisons of mentally manipulated objects). On the contrary CKAO, may not facilitate specific process components, but rather control the process components. In other words, CKAO may help a 'modeling' to graphicacy problems. In Chapter III, a similar attribution is made based on the mediating effects of content knowledge. Content knowledge mediates the relationship between time-on-task and comprehension success during the initial reading and during task completion phase in two different ways. Specifically, spending more time at the initial reading phase only leads to better comprehension when content knowledge is high. On the contrary, when content knowledge is low spending more time at initial reading does not lead to better comprehension. This mediation suggests that content knowledge influences text-graph comprehension because it controls the comprehension process. In other words, content knowledge provides the structure to build a coherent internal representation 'piece by piece' over time. With content knowledge, more time leads to a more coherent mental model because each 'piece' solidifies the coherence structure. Without content knowledge, more time does not lead to better coherence because the collection of the 'pieces' is unstructured. However, content knowledge influences comprehension in a different way during task completion. During task completion, the primary goal is not the construction of a coherent internal representation but to quickly find the task relevant information. Individuals with more content knowledge are both faster and more accurate during task completion. Individuals with low content knowledge only solved items when they invested more time, because it takes them more time to find the relevant information. The double sided mediation effect shows that content knowledge can facilitate comprehension based on two different underlying mechanisms.

Both Chapters (II and III) allow the attribution of specific influencing mechanisms. The CKAO and content knowledge (during initial reading) facilitate graphicacy performance because they control and structure the comprehension process, whereas arithmetic fluency and number line estimation, and content knowledge (during task completion) facilitate graphicacy performance because they facilitate specific process components. However, the double sided mediation effect of content knowledge shows that the functioning mechanism can change relative to the comprehension phase. The POMoG as it was proposed did not take the specific individual characteristics arithmetic fluency, number line estimation and CKAO into account, and did not consider different mechanisms relative to the comprehension phase. Therefore, the POMoG can be updated based in the presented studies. The updated POMoG considers the individual characteristics arithmetic fluency, number line estimation and CKAO. Furthermore, Chapter III showed that coherence-oriented processing influences comprehension success when content knowledge is high. Therefore, the updated POMoG considers a processing path for coherence-oriented and task selective processing. Coherence-oriented processing does not involve the internal representation of the task, in distinction to task-selective processing (See updated POMoG in figure 20).

Chapter IV demonstrated that an internal representation of the graph is not a prerequisite for an integrated comprehension of the text and graph. First of all, the results show that there are hierarchical dependencies between comprehension processes. These hierarchical dependencies determine which internal representation can be constructed after another. The POMoG it was assumed that an internal representation of the graph proceeds the internal representation of the content. However, the internal representation of the content can aids the construction of an internal representation of the graph, when text is present as another information source. This implies that the construction sequence of internal representations by the POMoG can deviate, whenever additional information sources are present. Therefore, the update POMoG includes the possibility of processing support the construction of the internal representation of the graph (See updated POMoG in figure 1).

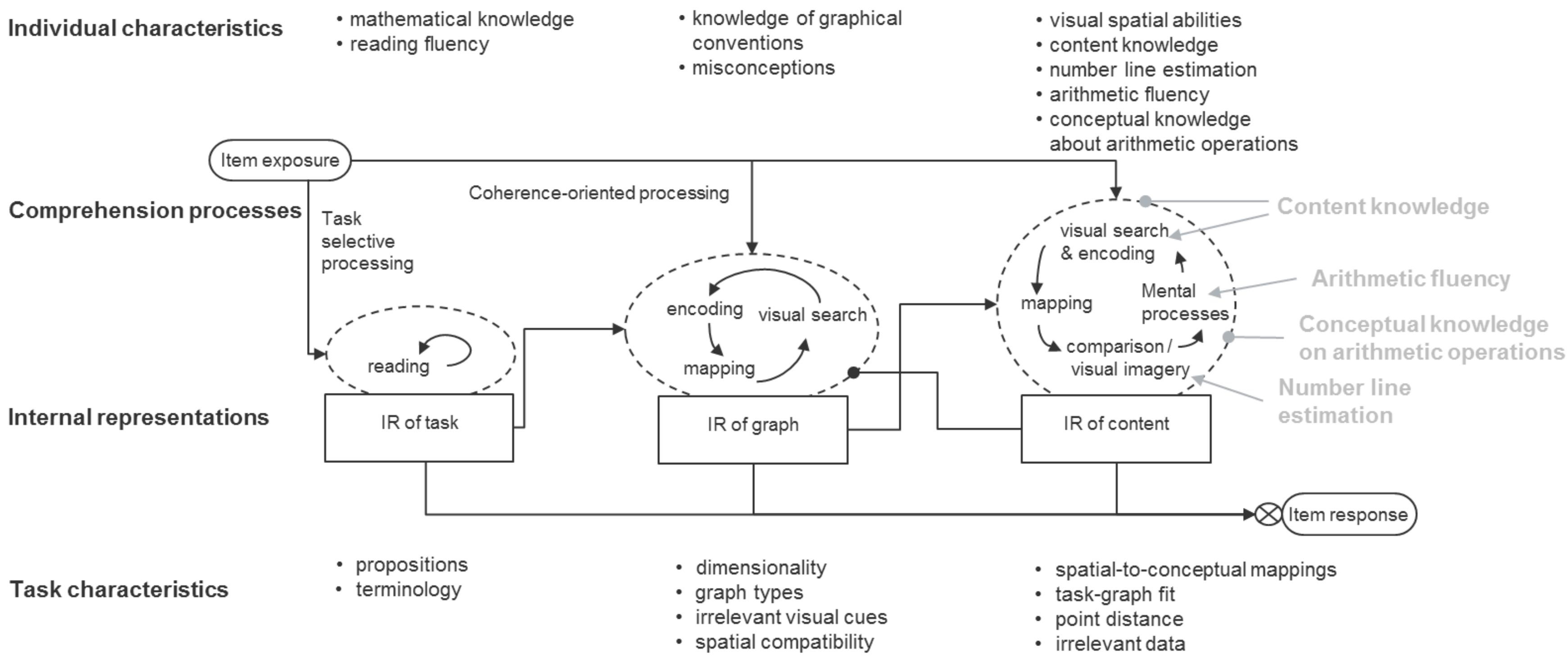


Figure 20. Updated Process-Oriented Model of Graphicacy including a novel coherence-oriented and task selective processing path and an augmented list of individual characteristics. In gray: Individual characteristics with influencing mechanism. Gray pointy arrow indicate influence on specific process competent. Gray round arrows indicate controlling or structuring function.

V.4 Methodological implications

The methodological implications evolve around the methodological integration of literacy research and comprehension research and the interpretation of process measures, especially gaze behavior can be interpreted when individuals have different test results.

The thesis laid out that the separation of graphicacy and graph comprehension research is partly different statistical methods and research designs. Traditionally, graphicacy research applied factor analysis (and IRT model) based on test data and graph comprehension applied analysis of variance based on experimental data. However, the thesis presented analysis approaches that integrate both perspectives and demonstrates that there is no fundamental difference between tests and experiments. A test is an experiment with repeated measurement. Each item is a measurement, whereas items characteristics are the experimental conditions. An experiment is a test. Every item of a test is an experimental condition. However, in research practice test and experiments are treated very differently. The test is usually constructed based on text book content or 'real world' problems (e.g., Åberg-Bengtsson & Ottosson, 2006; Ullrich et al., 2012). From the experimental perspective, item characteristics are not systematically manipulated, like an experimental condition. Subsequently, the unsystematic variation of items characteristics results in interpretations problems discussed in the first chapter (Section I.6.). On the other side, experimental studies usually assume that the experimental outcomes is one dimensional. From the factor analytical perspective, this practice could be problematic for measurement precision and potentially hind effects. Subsequently, both communities can befit from the perspective of the other. The statistical modeling approach presented in Section I.7 and applied in Chapter III integrates factor analytical and experimental perspectives. More general, the EMPPI is a Generalized Linear Mixed-Effect Regressions (GLMER; Bates, Mächler, Bolker, & Walker, 2014). GLMER allow researchers to treat experiments like a tests and tests like experiments (Baayen, Davidson & Bates, 2008). In sum, graphicacy research can befit from systematic item construction based on theoretical models (e.g., the POMoG). Graph comprehension research can befit from taking the dimensionality of experimental outcomes into account. Both communities can befit from using GLMERs to estimate the influence of item characteristic as an experimental conditions and to estimate the measurement precision of of experimental outcomes.

The second methodological implication concerns the eye-mind assumption (Carpenter & Just, 1975). The eye-mind assumption is the idea that gazes behavior 'represents the engagement of the mind'. Studies demonstrated that gaze behavior represents visual processing and attention allocation (e.g., Scheiter & Eitel, 2017), however, many studies are interest in mental processes (e.g., Hannus & Hyönä, 1999). However, mental process are not immediately to gaze behavior. Chapter III demonstrated that the link between text-graph transitions and information integration from text and graph is not immediate. Transitions between text and graph can be interpreted in two opposing ways relative to the comprehension phase. Text-graph transitions can indicates integration of information or disorientation, the inability to find relevant information. Therefore, the eye-mind assumption may be more accurate by stating that 'gaze behavior represents attention allocation and attention allocation allows under certain conditions inferences about the engagement of the mind'. In other words, the link between gaze behavior and mental processes is not immediate but, an indirect link can be established based on an experimental paradigm or a statistical model. In an experimental paradigm gaze behavior and mental processes can be linked because the stimuli is controlled and all can individuals master the experimental task. In this case, the gaze behavior represents the mental process because stimuli and outcome are the same, and difference in gaze behavior can be attributed to differences in mental processes. However, literacy research investigates individuals' differences. Consequently, the different outcomes are the research object. Gaze behavior cannot be linked to mental processes because individuals are more or less successful in different tasks. The POMoG uses an analogy to distance, time, and speed to illustrate this problem. The time it takes individuals to run a distance equals the individuals' speed, only if everyone runs the same distance. However, when the individuals run a different distances, their speed can be inferred by setting distance in relation to time. The same applies to the interpretation of gaze behavior, when individuals are more or less successful at performing the given tasks. The relationship between gaze behavior and individuals' success always inference about mental processes. The EMPPI is a statistical modeling approach that is based on this logic and allows inference about mental process relative to individuals' outcomes and gaze behavior. Chapter III demonstrated that the EMPPI can be applied to eye-tracking studies. Notably, to establish a strong link between gaze behavior and mental processes, the gaze behavior should be closely linked to the outcome. For instance, the number of text-graph transitions in one item should be linked to the item success in the same item. To tighten the link, process data needs to be analyzed on the response level, not aggregated across item which are partly solved and unsolved. In sum, gaze behavior can be linked to mental process based on experimental paradigms or statistical models. Experimental paradigms make gaze behavior interpretable by holding stimuli and outcome constant. Statistical modeling like the EMPPI enable inferences about mental processes by setting gaze behavior and outcomes in a relationship. The statistical modeling of gaze behavior and outcomes is necessary when individuals differ in their outcome. Therefore, studies that investigate mental process (i.e., not only attention allocation) based on gaze behavior, should either use an experimental paradigm in which ideally everyone solves every task, or use a statistical modeling approach that considers the relationship between task success and gaze behavior.

V.5 Practical implications

The practical implications evolve from the influencing mechanism of individual characteristics and the modeling approaches which could be used to create adaptive learning systems.

The thesis investigates the influence of various individuals characteristics on graphicacy and there influencing mechanisms. The thesis showed that BNAs influence graphicacy performance even in secondary education. Therefore, training BNAs, specifically arithmetic fluency, number line estimation, and CKAO may still be relevant in secondary education. Furthermore, it was argued that some individual characteristics either improve separate process components or help to structure and control process components. Training the controlling and structuring characteristics CKAO and content knowledge may be, especially effective at improving complex graphicacy performances. Therefore, graphicacy performance may be most effectively aided by improving students CKAO and content knowledge.

Chapter III showed a general positive association between text-graph transitions and comprehension success during initial reading, therefore, it may improve students' comprehension success by instructing students to perform more text-graph transitions during initial reading. In fact, interventions aiming at increasing transitions have been found to be effective at improving learning and comprehension (e.g., spatial contiguity: Johnson & Mayer, 2012; signaling: Ozcelik et al., 2009). However, the results also show a negative association between text-graph transition and the comprehension success in task completion. Therefore, the results suggest that it may be even detrimental for learning and comprehension to increase text-graph transitions by any means when students have to perform specific tasks. Increasing text-graph transitions can be detrimental because more transitions 'disrupt' the fluency of task-selective processing.

Furthermore, the modeling approaches applied in the thesis (i.e. KST & EMPPI) could be used to create adaptive systems learning systems (e.g., Aleven et at., in press), either based on individuals response pattern (Heller et al., 2006) or based on individuals' process measures (i.e. eye-movements; Schubert, 2016). The prerequisite relationships determined with the KST suggest an optimal comprehension paths (Heller et al., 2006) for comprehension of text and graph. The text-centered perspective implies relative from the comprehension process at which students struggle, instructional support can be more or less beneficial. For instance, the text-centered perspective implies that instructions aiming to improve text-graph comprehension would not be useful when text comprehension in itself is an issue. The interaction between process measure effect and individual characteristics implemented in the EMPPI could help to adjust adaptive systems to the individual needs of an individual.

Summary

The ability to understand visualizations of data has become immensely important for education, work, and life in the 21st century. In most cases, data is visualized as a graph. Graphs represent quantities via a 'paired with' relation. Graphs represent greater quantities by longer lines, higher bars, or more of some other visual dimension. Despite the presence of graphs in all areas of life, large-scale studies have raised concerns about students' ability to understand graphs. Compared to the long traditions of cognitive psychological work on reading and mathematics, relatively little is known about how individuals understand graphs. This thesis builds a cognitive psychological and psychometric model of the underlying comprehension processes, prerequisites and influential factors for individuals' ability to understand graphs.

A literature review revealed two separate research communities: first, literacy research describing individuals' ability to solve realistic problems with graphs (i.e., graphicacy research), and second, research explaining the underlying processes behind graph comprehension (i.e., graph comprehension research). In Chapter I, a Process-Oriented Model of Graphicacy (POMoG) was developed to integrate these research communities on a theoretical and methodological level. Chapters II, III and IV presented empirical studies that address different assumptions of the POMoG.

Chapter II investigated the influence of basic numerical abilities (BNAs) on graph reading performance. The influence of BNAs explains comprehension processes because specific BNAs can be linked to specific process components of graph reading. Subtraction, number line estimation, and conceptual knowledge about arithmetic operations were determined to be influencing factors based on test data from 750 students in secondary education. Subtraction and number line estimation facilitate unique process components, while conceptual knowledge helps students use efficient problem-solving strategies.

Chapter III investigated the association between time-on-task, text-graph transitions and comprehension success across comprehension phases during the comprehension of a text and graph. Text-graph transitions can either be interpreted as the integration of information or as disorientation, the inability to find relevant information. The association between time-on-task, text-graph transition and comprehension success was examined in two studies with 77 university students in total. The results showed that time-on-task and text-graph transitions are positively associated during the initial reading phase and negatively associated during the task completion phase. Text-graph transitions indicate integration during initial reading and disorientation during task completion. Additionally, students' content knowledge moderates the effect on time-on-task during initial reading and task completion. This moderating effect indicates that content knowledge can either control initial model construction or help students find task-relevant information.

Chapter IV investigated the dependency relationships between comprehension processes in the comprehension of a text and graph. Two contradicting perspectives about these relationships are present in the literature. The text-centered perspective states that comprehension of the graph depends upon comprehension of the text, and the multiple-representation perspective states that comprehension of the text and graph separately are both prerequisites for integrated comprehension. These perspectives were compared to the response patterns of 50 adults using knowledge space theory. The results showed that the text-centered perspective is more applicable to text-graph comprehension.

In sum, arithmetic fluency, number line estimation, conceptual knowledge about arithmetic operations, and content knowledge influence graphicacy via different underlying mechanisms. They either facilitate process components or help to control them. Further, it was demonstrated that process measures can be indicative of different comprehension processes depending on the comprehension phase. Finally, graph comprehension is not a prerequisite for integrated comprehension of a text and graph. The theoretical, methodological and practical implications of the thesis are discussed in Chapter V.

Zusammenfassung

Die Fähigkeit, Visualisierungen von Daten zu verstehen, ist für Bildung, Arbeit und Leben im 21. Jahrhundert enorm wichtig geworden. In den meisten Fällen werden die Daten als Graphen visualisiert. Graphen stellen Mengen über eine "gepaart mit"-Beziehung dar, wobei größere Mengen durch längere Linien, höhere Balken oder mehr von einer anderen visuellen Dimension dargestellt werden. Trotz des Vorhandenseins von Graphen in allen Lebensbereichen haben groß angelegte Studien Bedenken hinsichtlich der Fähigkeit von Schülern und Schülerinnen, Graphen zu verstehen, geäußert. Im Vergleich zu den langen Traditionen der kognitiven psychologischen Arbeit in den Bereichen Lesen und Mathematik ist relativ wenig darüber bekannt, wie Individuen Graphen verstehen. Diese Arbeit entwickelt ein kognitiv psychologisches und psychometrisches Modell, das die zugrunde liegenden Verständnisprozesse, Voraussetzungen und Einflussfaktoren der Fähigkeit, Graphen zu verstehen, abbildet. Im Kapitel I verwies die Literaturrecherche auf zwei getrennte Forschungsstränge. Zum einen die Literacy-Forschung, die die Fähigkeit des Einzelnen beschreibt, realistische Probleme mit Graphen zu lösen (sog. Graphicacy-Forschung), und zum anderen die Forschung, die die zugrunde liegenden Prozesse des Graphenverstehens erklärt (sog. Graphenverstehensforschung). Um diese Forschungsstränge auf theoretischer und methodischer Ebene zu integrieren, wurde ein prozessorientiertes Modell der Graphicacy (eng. POMoG) entwickelt. Empirische Studien, die sich mit den verschiedenen Annahmen des POMoG, wurden in den Kapiteln II, III und IV vorgestellt.

Kapitel II untersuchte den Einfluss von grundlegenden numerischen Fähigkeiten (eng. BNAs) auf die Graphenleseleistung. Der Einfluss von BNAs erklärt Verständnisprozesse, da die jeweiligen BNAs mit bestimmten Prozesskomponenten des Graphenlesens verknüpft werden können. Subtraktion, Zahlenstrahlschätzung und konzeptionelles Wissen über arithmetische Operationen wurden anhand von Testdaten von 750 Schülern aus der Sekundarstufe als Einflussfaktoren auf die Graphenleseleistung ermittelt. Subtraktion und Zahlenstrahlschätzung unterstützen einzelne Prozesskomponenten, während konzeptionelles Wissen bei der Anwendung effizienter Problemlösungsstrategien hilft.

Kapitel III untersuchte den Zusammenhang zwischen Bearbeitungszeit, Text-Graphen-Übergängen und Verständniserfolg über Verständnisphasen hinweg. Text-Graphen-Übergänge können entweder als Integration von Informationen oder als Desorientierung interpretiert werden. In zwei Studien mit insgesamt 77 Studierenden wurde der Zusammenhang zwischen Bearbeitungszeit, Text-Graphen-Übergänge und Verständniserfolg untersucht. Die Ergebnisse zeigten, dass Bearbeitungszeit und Text-Graphen-Übergänge während der initialen Lesephase positiv und während der Aufgabenerledigungsphase negativ assoziiert sein können. Beim initialen Lesen bedeuten mehr Text-Graphen-Übergänge mehr Integration,

während mehr Text-Graphen-Übergänge bei der Aufgabenbearbeitung auf Desorientierung hinweisen. Darüber hinaus moderiert das inhaltliche Wissen den Einfluss von Bearbeitungszeit auf den Verstehenserfolg während der initialen Lesephase und der Aufgabenerledigung. Dieser moderierende Effekt deutet darauf hin, dass inhaltliches Wissen entweder den anfänglichen Modellaufbau steuern oder den Schülern helfen kann, aufgabenrelevante Informationen zu finden.

Kapitel IV untersuchte die Abhängigkeitsbeziehung zwischen Verständnisprozessen bei Verstehen von Text und Graphen. Zwei widersprüchliche Perspektiven über diese Abhängigkeitsbeziehung sind in der Literatur vorhanden. Die textzentrierte Perspektive besagt, dass das Verständnis des Graphen vom Verständnis des Textes abhängt. Die Mehrfachrepräsentationsperspektive besagt, dass das Verständnis des Textes und des Graphen getrennt voneinander Voraussetzung für ein integriertes Verständnis ist. Diese Perspektiven wurden mit den Antwortmustern von 50 Erwachsenen mit Hilfe der Wissensraumtheorie verglichen. Die Ergebnisse zeigten, dass die textzentrierte Perspektive eher für das Verständnis von Text und Graphen geeignet ist.

Zusammenfassend lässt sich sagen, dass arithmetische Fähigkeiten, Zahlenstrahlschätzung, konzeptionelles Wissen über arithmetische Operationen sowie Inhaltswissen das Verstehen von Graphen über verschiedene zugrundeliegende Mechanismen beeinflusst. Entweder unterstützen sie Prozesskomponenten oder sie helfen bei deren Steuerung der Prozesskomponenten. Des Weiteren wurde gezeigt, dass Prozessmaßnahmen je nach Verständnisphase unterschiedliche Verständnisprozesse anzeigen können. Schließlich ist das Verständnis von Graphen keine Voraussetzung für ein integriertes Verständnis von Text und Graphen. Die theoretischen, methodischen und praktischen Implikationen der Arbeit wurden in Kapitel V diskutiert.

Declaration on Contributions to Monography

Although the dissertation is written as a monography, it includes contents of three manuscripts that are ready to submit. Their proportional contributions to the manuscripts are presented in the subsequent tables. This declaration and the tables can be found at the beginning of the dissertation as well.

Chapter II: Influences of basic numeric abilities on graph reading performance

Author	**Author position**	**Scientific ideas %**	**Data generation %**	**Analysis & interpretation %**	**Paper writing %**
Ulrich Ludewig	first	70	0	80	70
Katharina Lambert	second	0	50	0	0
Tanja Dackermann	third	0	50	0	0
Katharina Scheiter	fourth	15	0	0	10
Korbinian Moeller	fifth	15	0	20	20
Status in publication process:		Submitted			

Chapter III: Interpreting process measures in text-graphics comprehension

Author	**Author position**	**Scientific ideas %**	**Data generation %**	**Analysis & interpretation %**	**Paper writing %**
Ulrich Ludewig	first	70	100	70	60
Augustin Kelava	second	10	0	15	10
Korbinian Möller	third	0	0	0	10
Katharina Scheiter	fourth	20	0	15	20
Status in publication process:		Ready to submit			

Chapter IV: Contrasting a text-centered versus a multiple-representations perspective

Author	**Author position**	**Scientific ideas %**	**Data generation %**	**Analysis & interpretation %**	**Paper writing %**
Ulrich Ludewig	first	70	100	70	60
Augustin Kelava	second	15	0	15	10
Korbinian Moeller	third	0	0	0	10
Katharina Scheiter	fourth	15	0	15	20
Status in publication process:		Ready to submit			

Appendix Chapter II

Appendix Table 1. Results of multiple regression analysis and relative weights of basic numerical abilities, general cognitive ability, age, gender and two-way interaction between age, gender with basic numerical abilities and general cognitive ability.

	B	***β***	**[*L-CI,U-CI*]**	***RW***	***t***	***p***	***RS-RW* (%)**
Criteria = Graph reading performance							
[multiple R^2= .33, *adj.* R^2 = .31, $F(30,720) = 12.46$, p <.001]							
Intercept	6.71	.00	[-.06,.07]		80.53	.000	
Addition	-0.01	-.02	[-.14,.04]	.02	-0.35	.914	5.79*
Subtraction	0.09	.20	[.11,.29]	.05	4.17	.000	13.86*
Multiplication	0.01	.01	[-.05,.11]	.02	0.34	.914	6.22*
Number line estimation	2.13	.15	[.06,.19]	.04	4.07	.000	12.01*
Approximate arithmetic	0.01	.03	[-.04,.11]	.02	0.80	.748	6.12*
Conceptual knowledge	0.04	.11	[.03,.18]	.03	2.86	.026	9.92*
Basic geometry	0.01	.05	[-.03,.11]	.02	1.27	.616	5.17*
Non-sym. mag. comp.	0.01	.01	[-.04,.09]	.01	0.29	.925	1.70
G. cognitive ability	0.17	.30	[.22,.37]	.10	7.34	.000	28.54*
Gender[a]	0.07	.03	[-.04,.10]	.00	0.87	.748	1.16
Log(age)	0.14	.01	[-.06,.07]	.00	0.22	.954	0.46
Addition x age	0.2	.04	[-.05,.14]	.00	0.89	.748	1.31
Subtraction x age	-0.15	-.04	[-.13,.05]	.00	-0.82	.748	0.41
Multiplication x age	-0.01	.00	[-.09,.08]	.00	-0.06	.954	0.22
Number line estimation x age	2.56	.02	[-.05,.09]	.01	0.59	.828	2.26
Approximate arithmetic x age	-0.07	-.02	[-.09,.05]	.00	-0.54	.847	0.54
Conceptual knowledge x age	-0.11	-.03	[-.11,.04]	.00	-0.92	.748	0.34
Basic geometry x age	0.04	.04	[-.03,.11]	.00	1.12	.713	0.31
Non-sym. mag. comp. x age	0.18	.03	[-.04,.10]	.00	0.83	.748	0.61
G. cognitive ability x age	-0.11	-.03	[-.09,.05]	.00	-0.65	.828	0.80
Addition x gender	-0.06	-.11	[-.2, -.01]	.00	-2.16	.155	0.16
Subtraction x gender	0	.00	[-.09,.10]	.00	0.08	.954	0.12
Multiplication x gender	0.03	.06	[-.03,.13]	.00	1.33	.616	0.13
Number line estimation x gender	-1.08	-.07	[-.14,.00]	.00	-2.04	.178	0.09
Approximate arithmetic x gender	0	.00	[-.08,.09]	.00	0.07	.954	0.14
Conceptual knowledge x gender	-0.01	-.02	[-.09,.06]	.00	-0.46	.877	0.57
Basic geometry x gender	0	-.02	[-.10,.05]	.00	-0.63	.828	0.51
Non-sym. mag. comp. x gender	0.03	.05	[-.02,.11]	.00	1.29	.616	0.44
G. cognitive ability x gender	0	-.01	[-.08,.07]	.00	-0.13	.954	0.07
∑	-	-	-	.33	-	-	100.00

Note: *B*: unstandardized regression weight, *β:* standardized regression weight, *L-CI:* lower boundary of 95%-confidence interval, *L-CI:* upper boundary of 95%-confidence interval, *RW:* raw relative weight (within rounding error raw weights will sum to R²), *t* = t-value measures the size of the effect relative to the variation in sample data, *p:* FDR adjusted *p*-value, *RS-RW:* relative weight rescaled as a, percentage of predicted variance in the criterion variable attributed to each predictor (within rounding error rescaled weights sum to 100 %). [a] code female = -1, male = 1.* significantly different from a random variable.

Appendix Chapter III

Data Preparation study 1

Drift correction.We inspected all trials to check for plausibility and drift in eye movement patterns. We assumed that eye movement patterns for the text should be mostly sequential, while processing of the graph is more concentric. We applied a drift correction whenever fixations that were part of a sequential reading pattern fell into the graph area and fixations that were part of a concentric pattern fell into the text area. We used the first line of text to align the eye movement pattern with the stimulus material. We applied drift correction in 10% of all trials, and less than 1% of adjustments were larger than 150 pixels.

Tracking rate. As a next preparatory step in the analysis, we calculated a by-trial tracking rate defined as the ratio between tracked overall fixation duration and time-on-task. This tracking rate can be expected to be one hundred percent due to time for saccades and blinking. However, since we were analyzing the combination of time-on-task and eye movements, they should be proportional. We assumed that events that lower this tracking rate occur randomly. Therefore, we excluded trials in which fixation duration and time-on-task deviate to an extraordinary extent. We marked the eye-tracking measures for specific trials as missing when the ratio between tracked overall fixation duration and time-on-task was lower than 60%. A total of 68 trials did not reach this threshold.

Trimming process measures. Due to the amount of content for each item, we assumed that response times shorter than 5 seconds do not provide information about the cognitive processes of interest. Therefore, we excluded these cases. Since data collection took place in an experimental setting and all responses had a reasonable length, we did not define an upper boundary.

Data Preparation study 2. In the drift correction applied to 9% of trials, 8% of trials were deleted in the plausibility check. A further 205 trials had to be excluded due to a low tracking rate. This procedure led to the exclusion of 12 individuals from the analyzed sample. The remaining 88 invalid trials were spread evenly across individuals and items.

Again, two response times were below the 5-second threshold. Additionally, five initial reading times for the population dynamics material were much higher than average. The values are in line with events noted in the experiment protocol (a different browser version led to a difference in the display of the text bottom) and were marked as missing.

Appendix Table 2. Study 1: Means and Standard Deviations of Time-on-Task in Seconds and Count of Text-Graph Transitions during the Initial Reading Phase.

Material	**ToT** ***M* (*SD*)**	**TGT** ***M* (*SD*)**
Population dynamics	96.33 (9.99)	10.85 (0.56)
Action potentials	94.46 (8.15)	8.12 (1.05)
Sleep cycles	98.17 (9.49)	13.33 (1.41)
Overall	96.32 (3.06)	10.89 (0.35)

Note. N = 29. Means and standard deviations are pooled from five imputed datasets.

Appendix Table 3. Study 1: Means and Standard Deviations of Time-on-Task in Seconds and Count of Text-Graph Transitions during Task Completion Phase.

Material	**Number**	**Task**	**Accuracy**	**ToT** ***M* (*SD*)**	**TGT** ***M* (*SD*)**
Population dy-namics	1	2	.66	53.61 (38.60)	4.9 (0.92)
	2	1	.52	65.69 (50.35)	19.1 (5.92)
	3	1	.62	38.72 (16.75)	11.15 (1.28)
	4	1	.62	39.88 (35.62)	10.21 (2.36)
Action poten-tials	5	2	.66	56.12 (74.83)	5.13 (1.24)
	6	1	.55	56.71 (76.69)	16.49 (4.97)
	7	1	.17	59.42 (62.45)	18.29 (7.57)
	8	1	.59	58.28 (44.56)	13.33 (2.71)
Sleep cycles	9	2	.62	43.31 (15.88)	4.19 (0.69)
	10	2	.62	66.93 (74.31)	6.81 (1.48)
	11	1	.34	48.79 (43.04)	15.72 (4.47)
	12	1	.41	42.68 (16.86)	13.33 (1.99)
Overall			.53	52.51 (3.95)	11.56 (0.32)

Note. N = 29. Means and standard deviations are pooled from five imputed datasets.

Appendix Table 4. Study 1: Correlations and p-Values of Time-on-Task (ToT) and Text-Graph Transitions (TGT) during Initial Reading (IR) and Task Completion (TC).

	1	**2**	**3**	**4**
1. Accuracy	*r* (*p*)			
2.ToT:IR	.11 (.040)	*r* (*p*)		
3.TGT:TC	.10 (.067)	.58 (.000)	*r* (*p*)	
4.ToT:IR	-.17 (.001)	.12 (.022)	-.00 (.933)	*r* (*p*)
5.TGT:TC	-.24 (.000)	.07 (.229)	.02 (.766)	.67 (.000)

Note. N = 348. Correlations and *p*-values are pooled from five imputed datasets.

Appendix Table 5. Study 2: Means and Standard Deviations of Time-on-Task in Seconds and Count of Text-Graph Transitions during the Initial Reading Phase.

Material	**ToT** ***M*** **(*SD*)**	**TGT** ***M*** **(*SD*)**
Population dynamics	104.36 (4.74)	11.72 (0.51)
Action potentials	103.76 (5.63)	6.24 (0.15)
Sleep cycles	109.28 (7.76)	12.16 (0.57)
Overall	105.8 (2.02)	10.04 (0.15)

Note. N = 48. Means and standard deviations are pooled from five imputed datasets.

Appendix Table 6. Study 2: Means and Standard Deviations of Time-on-Task in Seconds and Count of Text-Graph Transitions during Task Completion Phase.

Material	**Number**	**Task**	**Accuracy**	**ToT** ***M*** **(*SD*)**	**TGT** ***M*** **(*SD*)**
Population dynamics	1	2	.54	58.48 (23.63)	5.98 (1.14)
	2	1	.40	60.45 (34.23)	15.47 (2.61)
	3	1	.48	43.66 (16.65)	12.24 (2.01)
	4	1	.67	42.91 (21.52)	11.43 (1.41)
Action potentials	5	2	.60	65.14 (47.87)	8.23 (2.77)
	6	1	.48	59.29 (34.19)	13.9 (2.53)
	7	1	.12	63.33 (28.59)	19.09 (3.04)
	8	1	.50	55.79 (27.77)	14.57 (2.52)
Sleep cycles	9	2	.75	54.62 (33.53)	6.8 (1.31)
	10	2	.69	60.35 (27.96)	7.59 (1.62)
	11	1	.42	58.43 (27.13)	18.3 (2.97)
	12	1	.52	46.56 (14.86)	14.57 (1.63)
Overall			.51	55.75 (2.39)	12.35 (0.21)

Note. N = 48. Means and standard deviations are pooled from five imputed datasets.

Appendix Table 7. Study 2: Correlations and p-Values of Time-on-Task (ToT) and Text-Graph Transitions (TGT) during Initial Reading (IR) and Task Completion (TC).

	1	**2**	**3**	**5**
1. Accuracy	*r* (*p*)			
2.ToT:IR	.01 (.745)	*r* (*p*)		
3.TGT:TC	.15 (.001)	.51 (.000)	*r* (*p*)	
4.ToT:IR	-.12 (.006)	.12 (.004)	-.06 (.180)	*r* (*p*)
5.TGT:TC	-.20 (.000)	.06 (.174)	-.04 (.509)	.69 (.000)

Note. N = 576. Correlations and *p*-values are pooled from multiple imputed datasets.

Appendix Table 8. Study 2: Means (M), Standard Deviations (SD), and Correlations of Individual Averages in Time-on-Task (ToT) and Text-Graph Transitions (TGT) during Initial Reading (IR) and Task Completion (TC), as well as Test Scores for Prior Knowledge (PK) and Graph (GC) and Reading Comprehension (RC).

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1. Acc								
2. Initial text view.	-.11 (.444)							
3. Initial graph view.	.24 (.093)	.19 (.198)						
4. Text viewing	-.23 (.111)	.47 (.001)	-.05 (.745)					
5. Graph viewing	0 (.996)	-.02 (.871)	.31 (.030)	.27 (.064)				
6. Prior knowledge	.35 (.014)	-.35 (.015)	.21 (.145)	-.46 (.001)	.03 (.842)			
7. Graph compr.	.38 (.007)	-.29 (.044)	.37 (.009)	-.24 (.095)	.35 (.016)	.47 (.001)		
8. Reading compr.	.22 (.136)	-.28 (.051)	.17 (.261)	-.26 (.075)	.02 (.880)	.37 (.010)	.27 (.066)	
M (SD)	0.51 (0.18)	75.23 (23.35)	10.02 (7.07)	21.09 (9.29)	22.48 (7.75)	18.42 (3.12)	17.15 (3.14)	11.02 (3.32)

Note. N = 48. Correlations, *p*-values, means, and standard deviations pooled from five imputed datasets.

Appendix Table 9. Study 2: Likelihood ratio test results for nested model comparison. model comparisons indicate whether alternative model is a better description of the data than the null model. the individual characteristics prior knowledge (PK) and graph (GC) and reading comprehension (RC) and their interactions with time-on-task (ToT) and text-graph transitions (TGT) during initial reading (IR) and task completion (TC) are subsequently added to the original model including the main effects of ToT and TGT during IR and TC.

Process measure	**Individual characteristic**	**Null model**	**Alternative model**	***df***	χ^2	***p***
ToT	PK	ToT	+ PK	1	5.25	**.022**
		ToT + PK	+ IR x PK	1	7.56	**.006**
		ToT + PK	+ TC x PK	1	10.29	**.001**
		ToT + PK	+ IR x PK + TC x PK	2	21.19	**<.001**
	GC	ToT	+ GC	1	12.58	**<.001**
		ToT + GC	+ IR x GC	1	2.23	.136
		ToT + GC	+ TC x GC	1	1.93	.165
	RC	TOT	+ RC	1	2.04	.153
TGT	PK	TGT	+ PK	1	5.26	**.022**
		TGT + PK	+ IR x PK	1	2.09	.148
		TGT + PK	+ TC x PK	1	0.02	1.0
	GC	TGT	+ GC	1	5.99	**.014**
		TGT + GC	+ IR x GC	1	0.50	.482
		TGT + GC	+ TC x GC	1	0.03	.863
	RC	TGT	+ RC	1	2.04	.153

Note. N = 576. χ^2 and *p*-values of nested model comparisons are combined from imputed datasets. Combination rules are based on Enders (2010, p. 239 ff. as cited by Robitzsch, Grund, & Henke, 2017). Bolded *p*-values < .05 indicate that the alternative model describes the data significantly better than the null model.

Appendix Chapter IV

Appendix Table 10. Shows the topics of the texts and the core concepts they contain.

Topics	**Concept**
(a) population dynamics	(1) predator peak after prey peak
	(2) peaks stay on the same level
	(3) predator mean below prey mean
(b) action potential	(4) threshold value -50mv
	(5) resting potential of -70mv
	(6) K^+ open after action potential peaks
(c) sleep cycle	(7) falling asleep in first non-REM sleep phase
	(8) four to six sleep cycles per night
	(9) deep sleep in the first half of the night

Appendix Formula 1: The AIC_c is calculated using the value of maximized log-likelihood function for the respective BLIM (Wagenmakers & Farrell, 2004). In principle, AIC_c relates the value of the maximized log-likelihood function to the number of parameters (k) of each model. The AIC_c additionally adds a constant to the AIC. This constant relates to the sample size (n) and the number of parameters (k):

$$AIC_c = 2k - 2\ln(logLike) + \frac{2k(k+1)}{n-k-1} \tag{4}$$

Appendix Formula 2: Akaike Weights (ω_i). The ω_i can be interpreted as conditional probabilities. The ω_i indicate which model is the most likely of all the hypothesized models.

$$\omega_i = \frac{\exp(-\frac{1}{2}\,\Delta_i)}{\sum_{i=1}^{m}\exp(-\frac{1}{2}\,\Delta_i)} \tag{5}$$

References

Åberg-Bengtsson, L. (1999). Dimensions of performance in the interpretation of diagrams, tables, and maps: Some gender differences in the Swedish Scholastic Aptitude Test. *Journal of Research in Science Teaching, 36,* 565-582. doi:10.1002/(SICI)1098-2736(199905)36:5<565::AID-TEA4>3.0.CO;2-L

Åberg-Bengtsson, L., & Ottosson, T. (2006). What lies behind graphicacy? Relating students' results on a test of graphically represented quantitative information to formal academic achievement. Journal of Research in Science Teaching, 43, 43-62.https://doi.org/10.1002/tea.20087

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, *16*, 183-198. https://doi.org/10.1016/j.learninstruc.2006.03.001

Ainsworth, S., & VanLabeke, N. (2004). Multiple forms of dynamic representation. *Learning and instruction*, *14*, 241-255. https://doi.org/10.1016/j.learninstruc.2004.06.002

Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. *The Journal of the Learning Sciences, 11*, 25-61.

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (in press). Instruction based on adaptive learning technologies. In R. E. Mayer & P. Alexander (Eds.), *Handbook of Research on Learning and Instruction*. Routledge.

Ananiadou, K., & Claro, M. (2009). *21st Century Skills and Competences for New Millennium Learners in OECD Countries* (Report No. 41). Paris: OECD Publishing. http://dx.doi.org/10.1787/218525261154

Arndt, J., Schüler, A., & Scheiter, K. (2015). Text–Picture Integration: How Delayed Testing Moderates Recognition of Pictorial Information in Multimedia Learning. *Applied Cognitive Psychology*, *29,* 702-712. https://doi.org/10.1002/acp.3154

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Bailey, D., Siegler, R. S., & Geary, D. C. (2014). Early predictors of middle school fraction knowledge. *Developmental Science, 17,* 775-785, doi: 10.1111/desc.12155.

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science*, *14*, 125-135. doi: 10.1111/j.1467-7687.2010.00962.x

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1-48. https://doi.org/10.18637/jss.v067.i01b

Baumert, J., Bos, W., & Watermann, R. (1998). *TIMSS/III: Schülerleistungen in Mathematik und den Naturwissenschaften am Ende der Sekundarstufe II im internationalen Vergleich: Zusammenfassung deskriptiver Ergebnisse.* Berlin: Max-Planck-Institut für

Bildungsforschung. Retrieved from: http://hdl.handle.net/11858/00-001M-0000-0025-A35F-4

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*, 289-300. Retrieved from http://www.jstor.org/stable/2346101

Berg, C. A., & Phillips, D. G. (1994). An investigation of the relationship between logical thinking structures and the ability to construct and interpret line graphs. *Journal of Research in Science Teaching*, *31*, 323-344. doi: 10.1002/tea.3660310404

Berg, C. A., & Smith, P. (1994). Assessing students' abilities to construct and interpret line graphs: Disparities between multiple-choice and free-response instruments. *Science Education, 78*, 527-554. doi:10.1002/sce.3730780602

Bieda, K. N., & Nathan, M. J. (2009). Representational disfluency in algebra: Evidence from student gestures and speech. *ZDM, 41*, 637-650. https://doi.org/10.1007/s11858-009-0198-0

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79*, 1016-1031. doi:10.1111/j.1467-8624.2008.01173.x

Brenner, M. E., Herman, S., Ho, H. Z., & Zimmer, J. M. (1999). Cross-national comparison of representational competence. *Journal for Research in Mathematics Education*, 541-557. Stable URL: https://www.jstor.org/stable/749773

Burnham, K. P., & Anderson, D. R. (2003). *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer Science & Business Media.

Butcher, K. (2014). The Multimedia Principle. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (Cambridge Handbooks in Psychology, pp. 174-205). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

Butterworth, B., Zorzi, M., Girelli, L., & Jonckheere, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *The Quarterly Journal of Experimental Psychology: Section A*, *54*, 1005-1029. doi: 10.1080/02724980143000064

Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*, 1-67. https://dx.doi.org/10.18637/jss.v045.i03

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction, 20*, 155-166. https://doi.org/10.1016/j.learninstruc.2009.02.014

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, *14*, 5-26. https://doi.org/10.1023/A:1013176309260

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: a psycholinguistic processing model of verification. *Psychological Review*, *82*, 45-73. http://dx.doi.org/10.1037/h0076248

Chomsky, N., & Halle, M. (1965). Some controversial questions in phonological theory. *Journal of Linguistics, 1*, 97-138. https://doi.org/10.1017/S0022226700001134

Cox, R., Romero, P., du Boulay, B., & Lutz, R. (2004, March). A cognitive processing perspective on student programmers'‘graphicacy’. In *International Conference on Theory and Application of Diagrams* (pp. 344-346). Springer, Berlin, Heidelberg.

Curcio, F. R. (1987). Comprehension of mathematical relationships expressed in graphs. *Journal for Research in Mathematics Education, 18*, 382-393. doi:10.2307/749086

Dackermann, T., Huber, S., Bahnmueller, J., Nuerk, H.-C., & Moeller, K. (2015). An integration of competing accounts on children's number line estimation. *Frontiers in Psychology, 6:884*. doi:10.3389/fpsyg.2015.00884

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., & Partchev, I. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software*, *39*, 1-28. https://www.jstatsoft.org/article/view/v039i12/v39i12.pdf

De La Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*, 179-199. Doi: 10.1007/s11336-011-9207-7

De Smedt, B., Noël, M. P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education, 2*, 48-55. doi:10.1016/j.tine.2013.06.001c

Dehaene, S. (2009). Origins of mathematical intuitions. *Annals of the New York Academy of Sciences*, *1156*, 232-259. doi:10.1111/j.1749-6632.2009.04469.x

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: Double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, *33*, 219-250. https://doi.org/10.1016/S0010-9452(08)70002-9

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive neuropsychology*, *20*, 487-506. doi:10.1080/02643290244000239

Deutsches, P. K., & Baumert, J. (Eds.). (2013). PISA 2000-Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland. Verlag: Springer-. doi:10.1007/978-3-322-97590-4

Doignon, J. P., & Falmagne, J. C. (1985). Spaces for the assessment of knowledge. *International journal of man-machine studies*, *23*, 175-196. Doi: 10.1016/S0020-7373(85)80031-6

Dowker, A. (2005). Individual differences in arithmetic: Implications for psychology, neuroscience and education. Psychology Press, Hove, New York, 2005. pp. 358. doi:10.1002/icd.531.

Eitel, A., Scheiter, K., Schüler, A., Nyström, M., & Holmqvist, K. (2013). How a picture facilitates the process of learning from text: Evidence for scaffolding. *Learning and Instruction*, *28*, 48-63. https://dx.doi.org/10.1016/j.learninstruc.2013.05.002

Embretson, S. E., & Daniel, R. C. (2008). Understanding and quantifying cognitive complexity level in mathematical problem solving items. *Psychology Science*, *50*, 328.

Falmagne, J. C., Koppen, M., Villano, M., Doignon, J. P., & Johannesen, L. (1990). Introduction to knowledge spaces How to build, test, and search them. *Psychological Review, 97,* 201-224. https://dx.doi.org/10.1037/0033-295X.97.2.201

Fischer, G. H. (1995). The linear logistic test model. In *Rasch models* (pp. 131-155). Springer, New York, NY. https://doi.org/10.1007/978-1-4612-4230-7_8

Fischer, M. H., & Shaki, S. (2014). Spatial associations in numerical cognition. From single digits to arithmetic. *The Quarterly Journal of Experimental Psychology*, *67*, 1461-1483. http://dx.doi.org/10.1080/17470218.2014.927515

Freedman, E. G., & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. In *Diagrammatic representation and inference* (pp. 18-30). Springer Berlin Heidelberg.

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education, 32*, 124-158. doi:10.2307/749671

Friendly, M. (2008). A brief history of data visualization. In *Handbook of data visualization* (pp. 15-56). Springer, Berlin, Heidelberg.

Galesic, M., & Garcia-Retamero, R. (2011). Graph literacy a cross-cultural comparison. *Medical Decision Making, 31*, 444-457. doi:10.1177/0272989X10373805

Geary, D. C., Frensch, P. A., & Wiley, J. G. (1993). Simple and complex mental subtraction: strategy choice and speed-of-processing differences in younger and older adults. *Psychology and aging, 8*, 242.

Gillan, D. J. (1995). Visual arithmetic, computational graphics, and the spatial metaphor. *Human Factors*, *37*, 766-780. https://doi.org/10.1518/001872095778995571

Gillan, D. J. (2009, October). A Componential model of human interaction with graphs: VII. A Review of the Mixed Arithmetic-Perceptual Model. In *Proceedings of the Human Factors and Ergonomics Society annual meeting* (Vol. 53, No. 12, pp. 829-833). Sage CA: Los Angeles, CA: SAGE Publications. doi: 0.1518/107118109X12524442637705

Gillan, D. J., & Lewis, R. (1994). A componential model of human interaction with graphs: 1. Linear regression modeling. *Human Factors*, *36*, 419-440. doi:10.1177/001872089403600303

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*, 608-626. http://dx.doi.org/10.1037/a0034716

Greiff, S., Wüstenberg, S., Molnár, G., Fischer, A., Funke, J., & Csapó, B. (2013). Complex problem solving in educational contexts—Something beyond g: Concept, assessment, measurement invariance, and construct validity. *Journal of Educational Psychology*, *105*, 364-379. http://dx.doi.org/10.1037/a0031856

Guthrie, J. T., Weber, S., & Kimmerly, N. (1993). Searching documents: Cognitive processes and deficits in understanding graphs, tables, and illustrations. *Contemporary Educational Psychology, 18*, 186-221. doi:10.1080/10862969709547949

Hannus, M., & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low-and high-ability children. *Contemporary educational psychology*, *24*, 95-123. Doi: 10.1006/ceps.1998.0987

Hartig, J., & Höhler, J. (2008). Representation of competencies in multidimensional IRT models with within-item and between-item multidimensionality. *Zeitschrift für Psychologie/Journal of Psychology*, *216*, 89-101.doi: 10.1027/0044-3409.216.2.89

Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. *Leistung und Leistungsdiagnostik*, 127-143. doi:10.1007/3-540-33020-8_9

Hegarty, M. (2013). Cognition, metacognition, and the design of maps. *Current Directions in Psychological Science*, *22*, 3-9. doi: 10.1177/0963721412469395

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, *32*, 717-742. https://doi.org/10.1006/jmla.1993.1036

Heller, J., & Wickelmaier, F. (2013). Minimum discrepancy estimation in probabilistic knowledge structures. *Electronic Notes in Discrete Mathematics, 42*, 49-56. https://dx.doi.org/10.1016/j.endm.2013.05.145

Heller, J., Steiner, C., Hockemeyer, C., & Albert, D. (2006). Competence-based knowledge structures for personalised learning. *International Journal on E-Learning, 5*, 75-88. Chesapeake, VA: Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/p/21759/.

Hill, M., Sharma, M. D., O'Byrne, J., & Airey, J. (2014). Developing and evaluating a survey for representational fluency in science. *International Journal of Innovation in Science and Mathematics Education (formerly CAL-laborate International)*, *22*. Retrieved from: https://openjournals.library.sydney.edu.au/index.php/CAL/article/download/7484/8465

Hollands, J. G., & Spence, I. (1998). Judging proportion with graphs: The summation model. *Applied Cognitive Psychology, 12*, 173-190. doi:10.1002/(SICI)1099-0720(199804)12:2<173::AID-ACP499>3.0.CO;2-K

Huestegge, L., & Philipp, A. M. (2011). Effects of spatial compatibility on integration processes in graph comprehension. *Attention, Perception, & Psychophysics*, *73*, 1903-1915.doi: 10.3758/s13414-011-0155-1

Hullman, J., Adar, E., & Shah, P. (2011). Benefitting infovis with visual difficulties. *IEEE Transactions on Visualization and Computer Graphics*, *17*, 2213-2222. doi: 10.1109/TVCG.2011.175

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction*, *20*, 172-176. http://dx.doi.org/10.1016/j.learninstruc.2009.02.013

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied, 18, 178-191.* http://dx.doi.org/10.1037/a0026923

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*, 1-19. doi:10.1207/S15327906MBR3501_1

Kintsch, W. (1988). The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*, *95*, 163-182. http://dx.doi.org/10.1037/0033-295X.95.2.163

Klieme, E., Hartig, J., & Rauch, D. (2008). The concept of competence in educational contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3-22). Ashland, OH, US: Hogrefe & Huber Publishers.

Kohl, P. B., & Finkelstein, N. D. (2005). Student representational competence and self-assessment when solving physics problems. *Physical Review Special Topics-Physics Education Research*, *1*, 1- 11. DOI:https://doi.org/10.1103/PhysRevSTPER.1.010104

Kolkman, M. E., Kroesbergen, E. H., & Leseman, P. P. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learning and Instruction, 25*, 95-103. doi:10.1016/j.learninstruc.2012.12.001

Korossy, K. (1999). Modeling knowledge as competence and performance. In D. Albert & J. Lukas (Eds.), *Knowledge spaces: Theories, empirical research, applications* (pp. 103-132). Mahwah, NJ: Lawrence Erlbaum.

Kosslyn, S. M. (1989). Understanding charts and graphics. *Applied Cognitive Psychology, 3*, 185-225. https://doi.org/10.1002/acp.2350030302

Kozma, R. B., & Russell, J. (1997). Multimedia and understanding: Expert and novice responses to different representations of chemical phenomena. *Journal of research in science teaching, 34*, 949-968. Doi: 10.1002/(SICI)1098-2736(199711)34:9<949::AID-TEA7>3.0.CO;2-U

Kozma, R., & Russell, J. (2005). Students becoming chemists: Developing representationl competence. In *Visualization in science education* (pp. 121-145). Springer, Dordrecht. https://doi.org/10.1007/1-4020-3613-2_8

Lachmayer, S. (2008). Entwicklung und Überprüfung eines Strukturmodells der Diagrammkompetenz für den Biologieunterricht (Doctoral dissertation) [Development and evaluation of a structural model of graphing competence in science education]. Retrieved from http://macau.uni-kiel.de/receive/dissertation_diss_00003041

Lai, K., Cabrera, J., Vitale, J. M., Madhok, J., Tinker, R., & Linn, M. C. (2016). Measuring graph comprehension, critique, and construction in science. *Journal of Science Education and Technology*, *25*, 665-681. doi: 10.1007/s10956-016-9621-9

Lai, M. L., Tsai, M. J., Yang, F. Y., Hsu, C. Y., Liu, T. C., Lee, S. W. Y., ... & Tsai, C. C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, *10*, 90-115. http://dx.doi.org/10.1016/j.edurev.2013.10.001

Leinhardt, G., Zaslavsky, O., & Stein, M. K. (1990). Functions, graphs, and graphing: Tasks, learning, and teaching. *Review of educational research*, *60*, 1-64. doi: 10.3102/00346543060001001

Leroi-Gourhan, A. (1988). *Hand und Wort: Die Evolution von Technik, Sprache und Kunst.* Frankfurt am Main: Suhrkamp.

Leuders, T., & Sodian, B. (2013). Inwiefern sind Kompetenzmodelle dazu geeignet kognitive Prozesse von Lernenden zu beschreiben?. *Zeitschrift für Erziehungswissenschaft*, *16*, 27-33. Doi: 10.1007/s11618-013-0381-5

Lindner, M. A., Eitel, A., Strobel, B., & Köller, O. (2017). Identifying processes underlying the multimedia effect in testing: An eye-movement analysis. *Learning and Instruction*, *47*, 91-102. https://doi.org/10.1016/j.learninstruc.2016.10.007

Link, T., Nuerk, H. C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic competencies. *The quarterly journal of experimental psychology*, *67*, 1597-1613. http://dx.doi.org/10.1080/17470218.2014.892517

Lowrie, T., & Diezmann, C. M. (2011). Solving graphics tasks: Gender differences in middle-school students. *Learning and Instruction, 21*, 109-125. doi:10.1016/j.learninstruc.2009.11.005

Lowrie, T., Diezmann, C. M., & Logan, T. (2012). A framework for mathematics graphical tasks: the influence of the graphic element on student sense making. *Mathematics Education Research Journal, 24,* 169-187. https://doi.org/10.1007/s13394-012-0036-5

Mason, L., Pluchino, P., Tornatora, M. C., & Ariasi, N. (2013). An eye-tracking study of learning from science text with concrete and abstract illustrations. *The Journal of Experimental Education, 81*, 356-384. https://doi.org/10.1080/00220973.2012.727885

Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, *60*, 95-109. https://doi.org/10.1016/j.compedu.2012.07.011

Mason, L., Tornatora, M. C., & Pluchino, P. (2015). Integrative processing of verbal and graphical information during re-reading predicts learning from illustrated text: an eye-movement study. *Reading and Writing*, *28*, 851-872. https://doi.org/10.1007/s11145-015-9552-5

Mayer, R. E. (2005). Cognitive Theory of Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 43-71). New York, NY, US: Cambridge University Press. http://dx.doi.org/10.1017/CBO9781139547369.005

McKenzie, D. L., & Padilla, M. J. (1986). The construction and validation of the test of graphing in science (TOGS). *Journal of Research in Science Teaching*, *23*, 571-579.doi: 10.1002/tea.3660230702

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation*, *51*, 297-384. https://doi.org/10.1016/S0079-7421(09)51009-2

Moeller, K., Pixner, S., Zuber, J., Kaufmann, L., & Nuerk, H. C. (2011). Early place-value understanding as a precursor for later arithmetic performance – A longitudinal study on numerical development. *Research in developmental disabilities, 32*, 1837-1851. doi:10.1016/j.ridd.2011.03.012

Morris, T. P., White I. R., & Royston, P. (2014) Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology, 14*, 75-87. https://doi.org/10.1186/1471-2288-14-75

Naumann, J., & Goldhammer, F. (2017). Time-on-task effects in digital reading are non-linear and moderated by persons' skills and tasks' demands. *Learning and Individual Differences*, *53*, 1-16. https://doi.org/10.1016/j.lindif.2016.10.002

Nitsch, R., & Bruder, R. (2014). Diagnoseinstrument zum Aufdecken von Lernschwierigkeiten im Bereich funktionaler Zusammenhänge. Universitätsbibliothek Dortmund.

Nitz, S., Ainsworth, S. E., Nerdel, C., & Prechtl, H. (2014). Do student perceptions of teaching predict the development of representational competence and biological knowledge?. *Learning and Instruction*, *31*, 13-22. Doi: 10.1016/j.learninstruc.2013.12.003

Nuerk, H. C., Moeller, K., Klein, E., Willmes, K., & Fischer, M. H. (2011). Extending the mental number line. *Zeitschrift Für Psychologie, 219*, 3-22. doi:10.1027/2151-2604/a000041

O'brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity, 41*, 673-690. doi:10.1007/s11135-006-9018-6

Ögren, M., Nyström, M., & Jarodzka, H. (2017). There's more to the multimedia effect than meets the eye: is seeing pictures believing?. *Instructional Science*, *45*, 263-287. https://doi.org/10.1007/s11251-016-9397-6

Okan, Y., Garcia-Retamero, R., Cokely, E. T., & Maldonado, A. (2012). Individual differences in graph literacy: Overcoming denominator neglect in risk comprehension. *Journal of Behavioral Decision Making*, *25*, 390-401. doi: 10.1177/0963721413491570

Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior*, *26*, 110-117. https://dx.doi.org/10.1016/j.chb.2009.09.001

Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Computers & Education*, *53*, 445-453. https://doi.org/10.1016/j.compedu.2009.03.002

Ozuru, Y., Dempsey, K., & McNamara, D. S. (2009). Prior knowledge, reading skill, and text cohesion in the comprehension of science texts. *Learning and Instruction*, *19*, 228-242. https://doi.org/10.1016/j.learninstruc.2008.04.003

Padilla, M. J., McKenzie, D. L., & Shaw, E. L. (1986). An examination of the line graphing ability of students in grades seven through twelve. *School Science and Mathematics*, *86*, 20-26. doi:10.1111/j.1949-8594.1986.tb11581.x

Pesenti, M., Thioux, M., Seron, X., & De Volder, A. (2000). Neuroanatomical substrates of Arabic number processing, numerical comparison, and simple addition: A PET study. *Journal of cognitive neuroscience*, *12*, 461-479. doi:10.1162/089892900562273

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), Artificial intelligence and the future of testing (pp. 73-126). Hillsdale, NJ: Erlbaum.

Prenzel, M., Gogolin, I., & Krüger, H. H. (Eds.). (2008). *Kompetenzdiagnostik: Zeitschrift für Erziehungswissenschaft. Sonderheft 8| 2007*. Springer-Verlag.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

R Development Core Team (2012). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review, 17*, 19-36. https://dx.doi.org/10.1016/j.edurev.2015.12.003

Richter, T., Isberner, M. B., Naumann, J., & Neeb, Y. (2013). Lexical quality and reading comprehension in primary school children. *Scientific Studies of Reading*, *17,* 415-434. http://dx.doi.org/10.1080/10888438.2013.764879

Robitzsch, A., Grund, S., & Henke, T. (2017). miceadds: Some additional multiple imputation functions, especially for mice. R package version 2. 9-15. https://CRAN.R-project.org/package=miceadds

Roth, W. M., & Bowen, G. M. (2001). Professionals read graphs: A semiotic analysis. *Journal for Research in mathematics Education*, 159-194. doi: 10.2307/749672

Royer, J. M., Hastings, C. N., & Hook, C. (1979). A sentence verification technique for measuring reading comprehension. *Journal of Literacy Research*, *11*, 355-363. https://doi.org/10.1080/10862967909547341

Rubin, D. B., & Schenker, N. (1986). Multiple imputation for interval estimation from simple random samples with ignorable nonresponse. *Journal of the American Statistical Association*, *81*, 366-374. https://www.tandfonline.com/doi/abs/10.1080/01621459.1986.10478280

Rupp, A. A., & Mislevy, R. J. (2007). Cognitive foundations of structured item response models. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 205-241). New York, NY, US: Cambridge University Press.http://dx.doi.org/10.1017/CBO9780511611186.008

Scheiter, K., & Eitel, A. (2015). Signals foster multimedia learning by supporting integration of highlighted text and diagram elements. *Learning and Instruction, 36*, 11-26. https://dx.doi.org/10.1016/j.learninstruc.2014.11.002

Scheiter, K., Schüler, A., Gerjets, P., Huk, T., & Hesse, F. W. (2014). Extending multimedia research: How do prerequisite knowledge and reading comprehension affect learning from text and pictures. *Computers in Human Behavior, 31*, 73-84. https://dx.doi.org/10.1016/j.chb.2013.09.022

Scheiter, S., Eitel, A. (2017). The Use of Eye Tracking as a Research and Instructional Tool in Multimedia Learning, *Eye-Tracking Technology Applications in Educational Research* (143-164). IGI Global. https://dx.doi.org/10.4018/978-1-5225-1005-5.ch008

Schmidgall, S. P., Eitel, A., & Scheiter, K. (in press). Why do learners who draw perform well? Investigating the role of visualization, generation and externalization in learner-generated drawing. *Learning and Instruction*. https://dx.doi.org/10.1016/j.learninstruc.2018.01.006

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental science*, *20*. doi: 10.1111/desc.12372

Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: Their interrelations in grades 5 and 6. *Journal of Educational Psychology, 101*, 359. doi:10.1037/a0013840

Schneider, W., Schlagmüller, M., & Ennemoser, M. (2007). *LGVT 6-12: Lesegeschwindigkeits- und -verständnistest für die Klassen 6-12*. Göttingen: Hogrefe.

Schnotz, W. (2005). An Integrated Model of Text and Picture Comprehension. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 71-103). New York, NY, US: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511816819.005

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), Cambridge handbook of multimedia learning (2nd ed., pp. 72–103). Cambridge: Cambridge University Press.

Schnotz, W., & Baadte, C. (2015). Surface and deep structures in graphics comprehension. *Memory & Cognition, 43*, 605-618. https://doi.org/10.3758/s13421-014-0490-2

Schnotz, W., & Wagner, I. (2018). Construction and Elaboration of Mental Models Through Strategic Conjoint Processing of Text and Pictures. *Journal of Educational Psychology*. http://dx.doi.org/10.1037/edu0000246

Schnotz, W., Ludewig, U., Ullrich, M., Horz, H., McElvany, N., & Baumert, J. (2014). Strategy shifts during learning from texts and pictures. *Journal of Educational Psychology*, *106*, 974-989. http://dx.doi.org/10.1037/a0037054

Schoenfeld, A. H. (1989). Explorations of students' mathematical beliefs and behavior. *Journal for research in mathematics education*, *20*, 338-355. doi: 10.2307/749440

Schroeders, U., Bucholtz, N., Formazin, M., & Wilhelm, O. (2013). Modality specificity of comprehension abilities in the sciences. *European Journal of Psychological Assessment*, *29*, 3 -11. Doi: 10.1027/1015-5759/a000114

Schubert, C. (2016). *Supporting Adequate Processing of Multimedia Instruction: Two Gaze-based Interventions* (Doctoral dissertation, Eberhard Karls Universität Tübingen). http://dx.doi.org/10.15496/publikation-13763

Schüler, A., Arndt, J., & Scheiter, K. (2015). Processing multimedia material: Does integration of text and pictures result in a single or two interconnected mental representations?. *Learning and Instruction*, *35*, 62-72. https://doi.org/10.1016/j.learninstruc.2014.09.005

Schwonke, R., Berthold, K., & Renkl, A. (2009). How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology*, *23*, 1227-1243. https://doi.org/10.1002/acp.1526

Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, *13*, 227-237. https://doi.org/10.1016/S0959-4752(02)00022-1

Shah, P., & Freedman, E. G. (2011). Bar and Line Graph Comprehension: An interaction of top-down and bottom-up processes. *Topics in Cognitive Science, 3*, 560-578. doi:10.1111/j.1756-8765.2009.01066.x

Shah, P., Freedman, E. G., & Vekiri, I. (2005). The Comprehension of Quantitative Information in Graphical Displays. In P. Shah (Ed.) & A. Miyake, *The Cambridge Handbook of Visuospatial Thinking* (pp. 426-476). New York, NY, US: Cambridge University Press.http://dx.doi.org/10.1017/CBO9780511610448.012

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of educational psychology*, *91*, 690 - 702. doi: 10.1037/0022-0663.91.4.690

Siegler, R. S. (2016). Magnitude knowledge: The common core of numerical development. *Developmental Science, 19*, 341-361. doi:10.1111/desc.12395

Stieff, M., Hegarty, M., & Deslongchamps, G. (2011). Identifying representational competence with multi-representational displays. *Cognition and Instruction*, *29*, 123-145. https://doi.org/10.1080/07370008.2010.507318

TIMSS 2011 Assessment (2013). *International Association for the Evaluation of Educational Achievement (IEA)*. Retrieved from https://nces.ed.gov/timss/pdf/TIMSS2011_G8_Math.pdf

Tonidandel, S., & LeBreton, J. M. (2015). RWA web: A free, comprehensive, web-based, and user-friendly tool for relative weight analyses. *Journal of Business and Psychology*, *30*, 207-216. doi:10.1007/s10869-014-9351-z

Tonidandel, S., LeBreton, J. M., & Johnson, J. W. (2009). Determining the statistical significance of relative weights. *Psychological Methods, 14,* 387-399. doi:10.1037/a0017735.

Trafton, J. G., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., & Raymond, P. D. (2000). Turning pictures into numbers: extracting and generating information from complex visualizations. *Int. J. Human-Computer Studies, 53*, 827-850. doi:10.1006/ijhc.2000.0419

Trickett, S. B., & Trafton, J. G. (2006, June). Toward a comprehensive model of graph comprehension: Making the case for spatial cognition. In *International Conference on Theory and Application of Diagrams* (pp. 286-300). Springer, Berlin, Heidelberg.

Ullrich, M., Schnotz, W., Horz, H., McElvany, N., Schroeder, S., & Baumert, J. (2012). Kognitionspsychologische Aspekte eines Kompetenzmodells zur Bild-Text-Integration. *Psychologische Rundschau*, *63*, 11–17. Doi: 10.1026/0033-3042/a000105

Van Der Linden, W. J. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement*, *46*, 247-272. https://doi.org/10.1111/j.1745-984.2009.00080.x

Verschaffel, L., De Corte, E., & Lasure, S. (1994). Realistic considerations in mathematical modeling of school arithmetic word problems. *Learning and Instruction, 4*, 273-294. doi:10.1016/0959-4752(94)90002-7

Von Aster, M. G., & Shalev, R. S. (2007). Number development and developmental dyscalculia. *Developmental Medicine & Child Neurology, 49*, 868-873. doi:10.1111/j.1469-8749.2007.00868.x

von Kotzebue, L., & Nerdel, C. (2015). Modellierung und Analyse des Professionswissens zur Diagrammkompetenz bei angehenden Biologielehrkräften. *Zeitschrift für Erziehungswissenschaft*, *18*, 687-712.doi: 10.1007/s11618-015-0639-1

Vukovic, R. K., Fuchs, L. S., Geary, D. C., Jordan, N. C., Gersten, R., & Siegler, R. S. (2014). Sources of individual differences in children's understanding of fractions. *Child development, 85*, 1461-1476. doi:10.1111/cdev.12218

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic bulletin & review*, *11*(1), 192-196. Doi: 10.3758/BF03206482

Wainer, H. (1992). Understanding graphs and tables. *Educational Researcher*, *21*, 14-23. doi:10.3102/0013189X021001014

Walker, C. M., & Beretvas, S. N. (2003). Comparing multidimensional and unidimensional proficiency classifications. multidimensional IRT as a diagnostic aid. *Journal of Educational Measurement*, *40*, 255–275. Retrieved from http://www.jstor.org/stable/1435130

Weiß, R., Albinus, B., & Arzt, D. (2006). Grundintelligenztest Skala 2-Revision (CFT 20-R). Hogrefe.

Wickelmaier, F., Helle, J., & Anselmi, P. (2016). psk: Probabilistic Knowledge Structures. R package version 0.4-0. https://CRAN.R-project.org/package=pks

Wilson, M., & De Boeck, P. (2004). Descriptive and explanatory item response models. In *Explanatory item response models*(pp. 43-74). Springer, New York, NY. doi: https://doi.org/10.1007/978-1-4757-3990-9_2

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, *14*, 97-116. doi: 10.1111/j.1745-3984.1977.tb00031.x

Wu, M., & Adams, R. (2007). *Applying the Rasch model to psycho-social measurement: A practical approach*. Melbourne: Educational Measurement Solutions.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*, 162 -185. https://dx.doi.org/10.1037/0033-2909.123.

List of tables

List of figures

Curriculum vitae

Ulrich Ludewig

ACADEMIC CAREER

10/2015 – present	Ph.D. candidate at the Graduate School & Research Netzwork 'LEAD' (Learning, Educational Achievement, and Life Course Development), funded by the German Excellence Initiative
10/2015 – present	Associate member of the Leibniz-Institut für Wissensmedien, Tübingen and member of the research group ‚multiple representations' (Prof. Dr. Katharina Scheiter)
04/2014 – 07/2014	Research intern in the project "Energy visualization in autonomous houses" in collaboration with BOUYGUES CONSTRUCTION and the Ideas Lab (Xavier Gauvin), Grenoble (supervised by Prof. Dr. Erica De Vries)
01/2013 – 06/2013	Market Research Intern at United Internet Media GmbH (supervised by Dr. Uli Gleich)
06/2012 – 12/2012	Research assistance in the project „BiTe – Integrative processes of pictures and texts in secondary education", (Prof. Dr. W. Schnotz), department of Educational Psychology at University Koblenz-Landau
02/2012 – 05/2012	Research intern in the project „BiTe – Integrative processes of pictures and texts in secondary education", (Prof. Dr. W. Schnotz), department of Educational Psychology at University Koblenz-Landau

EDUCATION

10/2013 – 10/2015	Master of Arts in International Cognitive Visualization, Interdisciplinary Studies, California State University, Chico (final grade: 1.2)
02/2014 – 07/2014	Erasmus Exchange Semester, University Pierre Mendès-France Grenoble, France
04/2010 – 09/2013	Bachelor of Science in Psychology, University Koblenz-Landau, Germany (final grade: 1.6)