

Fachgebiet: Psychologie

**TO CALIBRATE OR NOT TO CALIBRATE?  
CONDITIONS AND PROCESSES OF METACOGNITIVE  
CALIBRATION DURING HYPERMEDIA LEARNING**

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der

Philosophischen Fakultät

der

Westfälischen Wilhelms-Universität

zu

Münster (Westf.)

vorgelegt von

Stephanie Pieschl

aus Münster (Westf.)

Münster, 2007



Tag der mündlichen Prüfung: 06.02.2008  
Dekan: Prof. Dr. Woyke  
Referent: Prof. Dr. Bromme  
Korreferent: Prof. Dr. Winne



## Acknowledgements

Even though I am the person responsible for every written word in this Ph.D. thesis, any work is a communal effort. I got advice and support from many people, from colleagues as well as from friends and family. Thanks to all of you. However, some people will be addressed specifically because they contributed more to this thesis than others.

My dissertation was embedded in a project funded by the German Research Foundation (DFG) about the impact of epistemological beliefs on learning processes. The primary investigators of this project, Prof. Dr. Rainer Bromme and Prof. Dr. Elmar Stahl, were not only my superiors but also my main advisors for my thesis. Thank you for your support, your patience and many constructive discussions. Furthermore, student research assistants were involved in this project who not only contributed to this thesis by helping with data collection, but also by working independently on the project while I was preoccupied with my dissertation. Thank you: Stephanie Adrian, Sabine Joachim, Mike Pillukat, Torsten Porsch, Leoni Pytlik, and Janina Wolters. Furthermore, even though not all my colleagues from the research group of Prof. Dr. Rainer Bromme were directly involved in my thesis, they still contributed by being responsible for an enjoyable working climate. Thank you.

I owe further thanks to the faculty and the Ph.D. students of the Virtual Ph.D. Program (VGK) of the German Research Foundation (DFG) who gave constant feedback on my thesis. Especially, I want to thank Prof. Dr. Friedrich Hesse who was my advisor within this graduate program.

In addition, international researchers also contributed, especially Dr. Tom Murray who helped me set up MetaLinks and Prof. Dr. Phil Winne and his research team at Simon Fraser University who welcomed me during a research visit and gave valuable feedback.

Furthermore, I have to thank my friends for not abandoning me although I was often absent during the writing process. Last but not least, a big thank you to my family, especially to my parents Lothar and Sefi, but also to my sister Patricia and her husband Ilyes who worked on an even more important project during this time, my niece Menyar. I want to dedicate this thesis to my deceased grandparents, especially to Walter Pieschl who was about to become the first doctor in the family when World War II broke out.



<b>1</b>	<b>INTRODUCTION .....</b>	<b>1</b>
<b>2</b>	<b>THEORETICAL BACKGROUND .....</b>	<b>5</b>
<b>2.1</b>	<b>The COPES-Model of Studying (Winne &amp; Hadwin, 1998) .....</b>	<b>5</b>
2.1.1	The COPES Facets for Learning.....	7
2.1.2	The Four Stages of the COPES-Model.....	9
2.1.3	Metacognitive Processes in the COPES-Model .....	14
2.1.4	Conclusion.....	16
<b>2.2</b>	<b>Metacognitive Calibration .....</b>	<b>18</b>
2.2.1	Selective and Critical Review of Traditional Research on Calibration .....	19
2.2.2	An Extended Definition of the Construct Calibration.....	26
2.2.3	A New Research Strategy – Transfer of Methodology.....	28
2.2.4	Conclusion.....	34
<b>2.3</b>	<b>Task Complexity – An Important External Condition .....</b>	<b>35</b>
2.3.1	Theoretical Conceptualizations and Potential Operationalisations .....	35
2.3.2	Bloom’s (Revised) Taxonomy of Educational Objectives .....	39
2.3.3	Empirical Results and Corresponding Theoretical Explanations .....	48
2.3.4	Conclusion.....	50
<b>2.4</b>	<b>Text Complexity – An Important External Condition.....</b>	<b>51</b>
2.4.1	Theoretical Conceptualizations and Potential Operationalisations .....	52
2.4.2	Empirical Results and Corresponding Theoretical Explanations .....	54
2.4.3	Conclusion.....	56
<b>2.5</b>	<b>Prior Domain Knowledge – An Important Internal Condition .....</b>	<b>57</b>
2.5.1	Theoretical Framework – the Expert Paradigm .....	58
2.5.2	Empirical Results and Corresponding Theoretical Explanations .....	61
2.5.3	Conclusion.....	68
<b>2.6</b>	<b>Epistemological Beliefs – An Important Internal Condition .....</b>	<b>69</b>
2.6.1	Theoretical Conceptualizations and Potential Operationalisations .....	70
2.6.2	Empirical Results and Corresponding Theoretical Explanations .....	81
2.6.3	Conclusion.....	87
<b>3</b>	<b>EMPIRICAL STUDIES.....</b>	<b>91</b>
<b>3.1</b>	<b>Main Research Questions of the Empirical Studies .....</b>	<b>92</b>
3.1.1	Short Review of Research Questions.....	92
3.1.2	Operationalisation of Research Questions.....	94
<b>3.2</b>	<b>Hypertext on “Genetic Fingerprinting”.....</b>	<b>98</b>

3.2.1	MetaLinks – A Template for Hierarchical Hypertexts .....	99
3.2.2	Navigation in the Hypertext.....	100
3.2.3	The Overall Structure of the Hypertext on “Genetic Fingerprinting” .....	101
3.2.4	A Quantitative Description of the Hypertext on “Genetic Fingerprinting” .....	104
3.2.5	Chapter on mtDNA Analysis .....	106
3.2.6	Chapter on STR Analysis .....	108
3.2.7	Chapter on Y-STR Analysis .....	109
3.2.8	Formative and Summative Evaluation of Text Complexity .....	110
3.2.8.1	Pilot 1 – Formative Evaluation.....	110
3.2.8.2	Pilot 2 – Formative Evaluation.....	112
3.2.8.3	Pilot 3 – Summative Evaluation .....	113
3.2.8.4	Conclusion.....	115
<b>3.3</b>	<b>Study I - Calibration to Task Complexity in the Preparatory Stages.....</b>	<b>116</b>
3.3.1	Method.....	116
3.3.1.1	Procedure.....	116
3.3.1.2	Participants .....	117
3.3.1.3	Materials.....	118
3.3.2	Results.....	124
3.3.2.1	Do Students Discriminate Between Tasks of Different Complexity?.....	124
3.3.2.2	Do Students Calibrate to Task Complexity? .....	125
3.3.2.3	The Impact of Prior Domain Knowledge and Epistemological Beliefs .....	127
3.3.3	Local Discussion.....	130
3.3.3.1	Discrimination and Calibration.....	130
3.3.3.2	The Impact of Prior Domain Knowledge and Epistemological Beliefs .....	131
3.3.3.3	Open Issues and Implications.....	134
<b>3.4</b>	<b>Study II – Calibration to Task Complexity in the Enactment Stages .....</b>	<b>136</b>
3.4.1	Method.....	136
3.4.1.1	Procedure.....	136
3.4.1.2	Participants .....	137
3.4.1.3	Materials.....	138
3.4.2	Results.....	146
3.4.2.1	Do Students Discriminate Between Tasks of Different Complexity?.....	146
3.4.2.2	Do Students Calibrate to Task Complexity? .....	151
3.4.2.3	The Impact of Prior Domain Knowledge and Epistemological Beliefs .....	152
3.4.2.4	Determinants of the Learning Outcome .....	156
3.4.3	Local Discussion.....	161
3.4.3.1	Discrimination and Calibration.....	161
3.4.3.2	The Impact of Prior Domain Knowledge and Epistemological Beliefs .....	162
3.4.3.3	Determinants of the Learning Outcome .....	165
3.4.3.4	Open Issues and Implications.....	166
<b>3.5</b>	<b>Study III – Calibration to Text Complexity in the Enactment Stages .....</b>	<b>168</b>
3.5.1	Method.....	169
3.5.1.1	Procedure.....	169
3.5.1.2	Participants .....	169
3.5.1.3	Materials.....	170



3.5.2	Results .....	172
3.5.2.1	Do Students Discriminate Between Texts of Different Complexity? .....	175
3.5.2.2	Do Students Calibrate to Text Complexity?.....	175
3.5.2.3	The Impact of Prior Domain Knowledge and Epistemological Beliefs.....	176
3.5.2.4	Determinants of the Learning Outcome .....	181
3.5.3	Local Discussion.....	185
3.5.3.1	Discrimination and Calibration.....	185
3.5.3.2	The Impact of Prior Domain Knowledge and Epistemological Beliefs.....	186
3.5.3.3	Determinants of the Learning Outcome .....	189
3.5.3.4	Open Issues and Implications.....	190
<b>4</b>	<b>GENERAL DISCUSSION .....</b>	<b>193</b>
4.1	Are the Empirical Results Consistent with the COPES-Model?.....	194
4.2	Do Learners Adapt their Learning to External Conditions? .....	194
4.3	Which are the Most Important External Conditions?.....	199
4.4	Do Internal Conditions Impact these Adaptation Processes? .....	201
4.4.1	Prior Domain Knowledge .....	202
4.4.2	Epistemological Beliefs.....	205
4.5	Do Conditions Impact All Stages Alike? .....	209
4.6	Is More Flexible and Accurate Adaptation Beneficial for Learning? .....	211
4.7	Proposing Specifications of the COPES-Model.....	213
4.8	Practical Implications .....	216
4.9	Future Research.....	217
	<b>REFERENCES.....</b>	<b>221</b>
	<b>APPENDIX.....</b>	<b>253</b>



## 1 Introduction

“I have needed some time to complete this task. Therefore, this task was quite complex by itself, at least the manner it could have been completed or the style I would have preferred to complete it, demanding extensive background knowledge ...”

(student BJPM18, study II)

Humans are very adaptive organisms from an evolutionary perspective; however it is an open issue if this is also true – or even desirable – for everyday learning experiences. On the one hand theoretical models argue that adaptivity is a desirable quality in learners (chapter 2.1). From a logical point of view this is also obvious: Only if learners flexibly adjust their learning process to specific situations their strategies can adequately match external demands. For example, if learners are confronted with simple fact-finding tasks they should not engage in deep elaboration processes and if learners are confronted with more complex tasks requiring deep argumentation they should not enact superficial memorization strategies. On the other hand, empirical studies demonstrate that most learning processes are far from perfect, probably because learners do not adequately take external demands into account and do not adapt their learning process flexibly.

To systematically investigate this kind of adaptivity to external demands empirically, learning with hypermedia was selected as a setting for this thesis. This kind of self-regulated learning is ubiquitous and of growing importance. Living in today’s modern society requires constant knowledge acquisition in multiple domains. Because of the transient nature of scientific knowledge due to frequent new scientific discoveries (reduced “half life” of knowledge) most jobs require constant advanced training. Individuals also have to systematically gather and understand information to make informed decisions in their private lives. Since the advent of personal computers, these learning processes are often enacted in multimedia settings. Even though multimedia does not necessarily foster better learning (Clark & Feldon, 2005), good educational software offers beneficial features (e.g., giving feedback in lieu of a human tutor or providing additional information on demand). Additionally, this solitary setting permits focusing on each individual learner and hypermedia enable automatically collecting learning process variables (e.g., logfiles).

However, so far no systematic methodology to investigate this issue of adaptability in learning scenarios has been established yet: Empirical studies investigating learners’ adaptation have mostly just explored if students changed their strategies flexibly at all (without

systematically relating that change to external demands). Therefore, one crucial argument within this thesis is that learners' adaptivity with regard to external demands can be systematically investigated with a methodology transferred from the traditional calibration paradigm (chapter 2.2): If learning processes systematically vary in accordance with external demands, so-called calibration can be diagnosed. For example, a learner who enacts a simple search strategy for a simple fact-finding task and a deep elaboration strategy for a complex argumentation task can be considered well-calibrated. Note that this application required significant extensions of the original definition of the construct "calibration": In the traditional calibration paradigm calibration referred primarily to the match between students' metacognitive *judgments* and *their own* performance. In this thesis calibration also refers to the match between students' *whole learning process* and *external* demands. By these extensions more ecologically valid learning processes can be investigated and problems of the traditional calibration paradigm were transformed into interesting research questions. These *processes of metacognitive calibration* are one main focus of this thesis.

Based on a conceptual analysis of the specific affordances of hypermedia learning *task complexity* (chapter 2.3) and *text complexity* (chapter 2.4) emerged as potentially important external demands for this setting. More specifically, tasks were created that afforded differently complex cognitive processes (Anderson et al., 2001), for example simple tasks that required recall and complex tasks that required evaluation of content. To manipulate text complexity, a hypertext on "genetic fingerprinting" was constructed that encompasses texts of different complexity. This methodology permits investigating learners' calibration to these two specific external demands. Additionally, this methodology also permits investigating if this kind of adaptation is beneficial for learning ("*To calibrate or not to calibrate?*").

In addition to these most central issues, it is an open question if all learners are equally good at calibrating their learning processes to external demands or if their adaptations might depend on specific learner characteristics. If this question could be answered satisfactory, adequate instructional interventions could be devised to foster better learning and superior adaptation. Most theoretical models as well as most empirical results point to the crucial role of *prior domain knowledge* (chapter 2.5) for learning: In the scenario of this thesis, prior domain knowledge might allow for a more accurate diagnosis of task complexity or text complexity and thus might lead to better calibration. However, the role of other learner characteristics is less clear. For example, most theories consider learners' beliefs about knowledge and knowing (*epistemological beliefs*, chapter 2.6) very important, but empirical results are inconsistent. In the scenario of this thesis, beliefs in the complexity and uncertainty of knowledge might trigger more elaborate learning processes and potentially better adaptation. The complex interactions between these two learner characteristics and the external demands task complexity and text complexity will be scrutinized in this thesis: In other words, not only the metacognitive calibration *processes*, but also important *conditions* for calibration will be investigated (*processes and conditions of metacognitive calibration*).

*Empirical Studies Devised to Investigate these Questions in Detail*

The COPES-model of self-regulated learning (Winne & Hadwin, 1998; chapter 2.1), supplemented by the theoretical ideas and the methodology transferred from the traditional calibration paradigm (chapter 2.2), served as a heuristic for the design of a series of empirical studies explicitly addressing these issues: The *conditions and processes of metacognitive calibration during hypermedia learning* will be scrutinized and their benefit for learning will be determined (“*To calibrate or not to calibrate?*”).

Study I exclusively focuses on the first preparatory planning stages of learning. Students are presented with six tasks of different complexity and have to give judgments (e.g., about the importance of learning strategies like memorizing) for each task. Based on these results it will be possible to determine if students are able to monitor task complexity and if they possess the necessary metacognitive knowledge to flexibly plan adequate strategies for differently complex tasks (*calibration to task complexity in the preparatory stages of learning*).

Study II primarily focuses on the subsequent enactment stages of learning. Students have to actively solve tasks of different complexity with a hypertext on “genetic fingerprinting”. Simultaneously, their concurrent thoughts and their concurrent actions are captured. Based on these results it will be possible to determine if students are also able to flexibly enact adequate learning strategies for differently complex tasks and if better adaptation is associated with better learning outcome (*calibration to task complexity in the enactment stages of learning*).

Study III also focuses on the enactment stages of learning, but with a different focus: text complexity. Students have to “learn as much as possible” about mtDNA analysis with a hypertext chapter with differently complex hierarchical levels. Simultaneously, their concurrent actions and their judgments regarding text comprehensibility are captured. Based on these results, it will be possible to determine if students are able to flexibly enact adequate learning strategies for differently complex texts and if better adaptation is associated with better learning outcome (*calibration to text complexity in the enactment stages of learning*).

*Theoretical and Practical Contributions of this Thesis*

On a theoretical level, this thesis is grounded in theories of self-regulated learning. More specifically, the COPES-model of self-regulated learning (Winne & Hadwin, 1998) is utilized as a general framework. The empirical results of this thesis will be able to validate or refute the core assumptions of this model. But more importantly, the empirical results will enable further specifications of models of self-regulated learning with regard to the complex interactions between external demands, internal learner characteristics, and the learning process. For example, the following questions can be answered: Do learners adapt their learning to external demands? Which are the most important external demands, task complexity or text complexity? How exactly do prior domain knowledge and epistemological beliefs impact these adaptation processes? Do these learner characteristics impact all stages

of learning alike and do they interact similarly with all external demands? And is more flexible adaptation to external demands really beneficial for the learning outcome?

Additionally, the traditional calibration paradigm is utilized as a more specific theoretical background. A detailed theoretical conceptualization as well as a methodology was transferred to a new application context: Students' adaptation processes to external demands are re-conceptualized as metacognitive calibration processes. The empirical results of this thesis will show if this transfer is successful and if this methodology constitutes a valuable addition to the methodological repertoire to investigate adaptation. Right now, the benefit of this methodology can be assumed, because so far students' adaptive use of learning strategies was mostly investigated without adequately taking into account external demands: For example, it was investigated as "fluctuation" (Moos & Azevedo, 2006) or as calibration to learners' internal standards (Luwel, Verschaffel, Onghena, & De Corte, 2003).

Furthermore, the empirical results of this thesis will facilitate the development of new instructional interventions to foster better self-regulated learning on a practical level. Because this is also the goal of a larger research project in which this thesis is embedded, one instructional intervention is already tested in the empirical studies of this thesis: An epistemological sensitization is administered to foster more "sophisticated" epistemological beliefs which in turn should lead to deeper elaboration learning processes (study I and study II). Empirical results will show the potential benefits of this strategy. Furthermore, detangling the complex interactions between external demands (task complexity and text complexity), learner characteristics (prior domain knowledge and epistemological beliefs), and self-regulated learning in different stages (preparatory stage and enactment stage) will help to better tailor scaffolds to student-specific deficits. For example, if empirical results show metacognitive knowledge deficits in the preparatory stages then it would not be helpful to implement scaffolds for strategy enactment.

## 2 Theoretical Background

“I think that this knowledge is rather complex as many pieces of information are intertwined with one another and if you want to achieve an overview, you have to really deeply elaborate the information, only a few minutes of text reading is not enough.”

(student AALF23, study II)

The COPES-model (Winne & Hadwin, 1998) will be used as theoretical framework for the series of empirical studies within this thesis (chapter 3), that is for the derivation of research questions as well as for the design of the studies. Consequently, it will also be utilized as a framework for structuring this theoretical background chapter.

The first sub-chapter (chapter 2.1) will describe the COPES-model in detail. Learners' adaptivity to *external conditions* and the impact of *internal conditions* on this process will be elaborated as the core research questions of this thesis. The second sub-chapter (chapter 2.2) will argue that these adaptation processes could be empirically investigated on a more detailed level with the methodology transferred from the *traditional calibration paradigm*. Consistently, two extensions of the traditional construct definition of “calibration” will be proposed that will also alleviate problems highlighted in the critical review of this paradigm. The subsequent sub-chapters will scrutinize the external conditions *task complexity* (chapter 2.3) and *text complexity* (chapter 2.4) that learners may adapt to. The last sub-chapters will detail *prior domain knowledge* (chapter 2.5) and *epistemological beliefs* (chapter 2.6) as potentially important internal conditions that might impact these adaptation processes. For external as well as for internal conditions adequate operationalisations will be proposed and empirical findings will be reviewed indicating which kinds of effects could be predicted.

### 2.1 The COPES-Model of Studying (Winne & Hadwin, 1998)

The COPES-model was proposed by Winne and colleagues (for a detailed description of the model see Winne & Hadwin, 1998<sup>1</sup>, Table 1 and Figure 2.1-1; for previous versions see Butler & Winne, 1995; Winne, 1995, 1996, 1997; and for subsequent elaborations see Hadwin & Winne, 2001; Winne, 2001; Winne, Jamieson-Noel, & Muis, 2002; Winne & Perry,

---

<sup>1</sup> All descriptions in this chapter are based on this most detailed description of the COPES-model unless otherwise specified.

2000). To give a short preview: With a strong focus on metacognitive processes, especially on *metacognitive monitoring*, the COPES-model assumes four stages of studying: (1) *task definition*, (2) *goal setting and planning*, (3) *enactment* of tactics and strategies, and (4) *adaptation*. All stages supposedly rely on the same facets of learning: *conditions*, *operations*, *products*, *evaluations*, and *standards* (COPES). Based on these constituents, the model describes how learners metacognitively adapt their studying to *external task conditions* and assumes a significant impact of *internal learner conditions*. The subsequent chapters contain a detailed description of these core elements of the COPES-model.

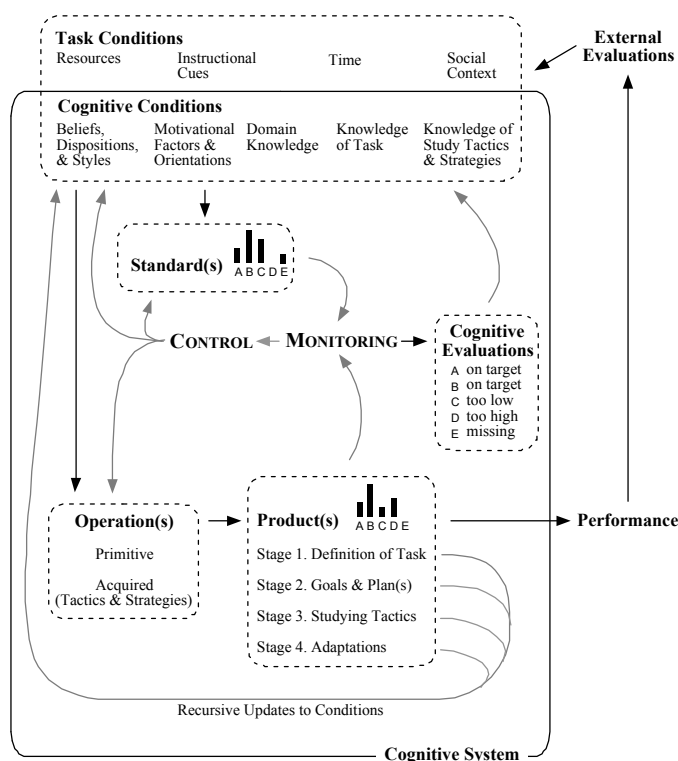


Figure 2.1-1: The COPES-model (Figure taken from Winne & Hadwin, 1998, p. 282)

The COPES-model was distilled from “a spectrum of educational and psychological models” (Butler & Winne, 1995, p. 247). It draws from heterogeneous theories about cognition, motivation, metacognition and especially information processing (Winne, 2001; Winne & Hadwin, 2001), but focuses on “cognition primus inter pares” (Winne, 2005, p. 233). It is important to note that the COPES-model’s primary focus is studying, that is inspecting information under the goal of learning (Winne, Jamieson-Noel, & Muis, 2002). *Studying* is assumed to differ from *learning* in multiple aspects: Studying rarely includes any scaffolding, it is most often a solo activity, it often involves searching in and synthesizing



multiple information sources, it often occurs in settings where the student can engineer her<sup>2</sup> own environment and it almost always produces observable traces (Winne & Hadwin, 1998). Furthermore, the COPES-model presumes that self-regulation is inherent to all learning activities (Winne, 1995) and develops through the learner's own experimentation with tactics and strategies (Winne, 1997).

### 2.1.1 The COPES Facets for Learning

First, the facets for learning will be detailed because all other core elements (see subsequent chapters) rely on these facets and because these features are unique to the COPES-model (Winne & Hadwin, 1998). A five facet schema is used to characterize instances of studying or studying tasks in terms of common dimensions: *conditions*, *operations*, *products*, *evaluations* and *standards*. This typology is referred to by the first-letter acronym COPES – and thus gave the model its name – because it identifies facets of tasks that a student has to “cope with” in studying and learning. Each of the slots in this script has to be filled for a specific task in a specific domain. More specifically, Winne (1997) argues that any studying activity requires three kinds of cognitive resources: (1) the *condition* slot (beliefs about under which conditions a tactic's product can be useful), (2) the *operation* slot (tactics that manipulate information), and (3) the *product* slot (understanding which tactic results in which kind of product). Furthermore, to accommodate *metacognitive monitoring* processes a learner would require some kind of internal (4) *standards* against which her behavior could be monitored or mapped: “Without a standard, monitoring is impossible” (Winne, 1997, p. 400). An additional (5) *evaluation* slot is needed to record the discrepancies between learners' standards and their behavior. These five facets are assumed to “interact with one another over time and, most likely, in a hierarchical or cascading manner” (Winne, Jamieson-Noel, & Muis, 2002, p. 126; Figure 2.1-1).

#### *Conditions*

Conditions represent the environment of the studying task and can be either external task conditions or internal cognitive conditions. Social context, time for task completion, instructional cues or resources such as books or the internet constitute *external* task conditions. Furthermore, *task complexity* (chapter 2.3) and the complexity of the material-to-learn (*text complexity*, chapter 2.4) can also be considered external task conditions. All kinds of learner variables can be subsumed under the roof of *internal cognitive conditions*. Butler and

---

<sup>2</sup> A coin was flipped to assign a gender to this hypothetical student (and all subsequently described individuals throughout this thesis). Although I will constantly refer to all students as female, this was done rather to improve readability than to indicate that a specific behavior is typical of females. Instead, the described students should represent both genders equally.

Winne (1995), for example, explicitly considered task knowledge (Winne & Marx, 1982; Graham, Schwartz, & MacArthur, 1993), strategy knowledge (Alexander & Judy, 1988; Borkowski, Chan, & Muthukrishna, 2000), self-efficacy (Bandura, 1989; Zimmerman, 1989), *prior domain knowledge* (chapter 2.5, Alexander & Judy, 1988; Chinn & Brewer, 1993), and *epistemological beliefs* (chapter 2.6, Schommer, 1990). All task conditions (internal and external) are assumed to act as a filter through which the studying task is approached in all stages of learning; thus, depending on specific conditions learners might derive different task definitions, goals and plans, enactments and adaptations (see stages of studying).

### *Operations*

Operations can be interpreted as specific kinds of schemas or information chunks that take the form of IF-THEN action chains and manipulate information. *Tactics* are interpreted as single IF-THEN rules. For example, such a rule for note taking could state: IF a definition in the text is found, THEN make a note of that definition. Most learning strategies such as underlining, summarizing or concept-mapping, can be considered tactics. A *strategy* on the other hand is a more comprehensive plan of how to tackle a task to achieve a specific goal and usually incorporates multiple tactics or a sequence of tactics in the form of IF-THEN-ELSE rules. The PQ4R strategy of previewing (P), questioning (Q), reading (R), reflecting (R), reciting (R), and reviewing (R) (Thomas & Robinson, 1982) might be considered a strategy. Based on theories of information processing (Anderson, 1991) Winne (2001) proposed five fundamental types of operations: First, to actively retrieve any information from long-term memory the learner needs to *search* her memory. Second, within this search process the learner has to *monitor* the products of the spreading activation for the relevant piece of knowledge she wants to retrieve. Third, to learn something (i.e., to add a chunk of information to long-term memory), new links need to be *assembled* by the learner. Fourth, to strengthen these links within memory this information has to be *rehearsed*. Fifth, as different formats of information are assumed be part of memory (e.g., visual vs. auditory) the learner must be able to *translate* one format into another. These strategies can be referred to by the acronym SMART operations (*searching, monitoring, assembling, rehearsing, and translating*). Note that not only cognitive tactics and strategies can be considered operations, but also motivational and emotional control strategies. Additionally, it is important to keep in mind that the COPES-model does not assume that all operations have to be deliberately regulated, rather for “familiar tasks actions may proceed automatically” (Butler & Winne, 1995, p. 258), probably because the whole strategy is melted into one chunk (Winne, 2001).

### *Products*

All operations create internal or external products. All kinds of *internal* products are transformed conditions for subsequent stages of studying or subsequent learning tasks (see paragraph about adaptation). For example, a newly acquired piece of knowledge would

constitute a change in domain knowledge, which in turn would act as condition for subsequent stages of studying or new studying tasks. *External* products on the other hand can usually be observed as study traces (e.g., observable behavior, internet logs, underlined text, written notes or essays, drawn concept maps, or written answers to studying questions). Each product – internal as well as external – is assumed to be represented as a multivariate profile of attributes. For example, a newly learned definition (internal product) could possess the attributes “hard to memorize”, “highly interrelated with other concepts”, “very abstract” and “very important and central”.

### *Standards*

Studying is considered to be a goal-directed behavior within the COPES-model (see paragraph about goal setting and planning). A goal can be represented as a multivariate profile of standards which characterize ideal, optimal or satisfying states. For example, standards might include learners’ goals with regard to targeted time-on-task, targeted level of understanding, or targeted rate of progress. Goals on other dimensions are feasible. Elaborating the example of the newly learned definition, one learner may hold the rather superficial goal (standard) to memorize definitions “very quickly”. A second learner may hold more elaborate goals (standards), for example “to memorize most important concepts” and “to find applications for each concept”.

### *Evaluations*

Learners are assumed to exercise constant metacognitive monitoring (see paragraph about metacognitive monitoring) by which they compares the product’s profile of attributes (see paragraph about products) against the profile of standards (see previous paragraph about standards) and thus generate a third multivariate profile of evaluations. These evaluations denote the congruencies and discrepancies between the two previous profiles. They mark whether an attribute is on target, off target, how much off target, or if the lists of attributes match at all. To illustrate this process, recall the above mentioned example of the learned definition. The first learner would notice at least two discrepancies: first, on the attribute “speed of memorizing” the product (“hard to memorize”) and the standard (“memorize very quickly”) differ and second, the number and kind of recorded attributes differ. Sometimes, these *internal* evaluations can be supplemented by *external* feedback created by performances (e.g., error messages from software applications or comments from teachers).

## **2.1.2 The Four Stages of the COPES-Model**

Most models of self-regulated learning differentiate at least some kind of *preparatory* phase, some kind of *performance* phase and some kind of *appraisal* phase (Boekaerts, 1992, 1995,

1996; Borkowski & Burke, 1996; Borkowski, Chan, & Muthukrishna, 2000; Pressley & Ghatala, 1990; Pintrich, 2000; Zimmerman, 2002; for reviews see Montalvo & Torres, 2004; Puustinen & Pulkkinen, 2001). This notion is consistent with the proposed stages of the COPES-model (Winne & Hadwin, 1998): Studying is assumed to progress through four distinguishable but recursively linked and weakly sequenced stages: (1) *task definition*, (2) *goal setting and planning*, (3) *enactment*, and (4) *adaptation*. The term “recursive” is used to describe that “products of earlier stages (or steps within a stage) update conditions on which operations work during the next cycle of activities” (Winne & Hadwin, 1998, p. 281). For example, information generated in the *enactment* phase may feed back into the same phase if the learner re-cycles through that phase or products of monitoring in later phases such as *enactment* might feed back to prior phases such as *task definition* if the learner re-cycles to earlier stages of learning (Winne & Perry, 2000; Winne, 2001). The term “weakly sequenced” refers to the fact that although most often the four stages unfold in the predefined order from stage one through stage four, there may be exceptions. For example, some tasks may be so familiar that the first stage is virtually skipped. Furthermore, the optional *adaptation* stage can basically be entered any time or it can be thoroughly skipped. All stages are assumed to rely on the same cognitive architecture or facets of learning (see previous chapter), but are differentiated by the products they generate (see Figure 2.1-1 and Table 1).

#### *Task Definition – The First Stage of Studying*

In the *task definition* stage of the COPES-model (Winne & Hadwin, 1998, see Table 1) a perception of the study task as a study space is generated. This kind of task definition entails the learners’ perceptions about possible constraints (e.g., limited time), available resources (e.g., library, internet or knowledgeable peers) and the given goal of the task. Thus, the product of this stage of studying is a comprehensive perception of the task in all its COPES facets (*conditions, operations, products, evaluations, and standards*).

As most tasks are not externally specified in all their facets, task perception may vary tremendously between learners. Winne (1997) assumes that a learner constructs a complete task representation based on a mix of bottom-up processing (data from the task environment, *external conditions*) and top-down processing (prior domain knowledge or prior knowledge about tasks, *internal conditions*). To give an example for the impact of external conditions: if a student is presented with the task to “learn” words, the targeted standard is still ambiguous. While it could mean to prepare for verbatim recall, it could also mean to prepare for stimulated recall or to be able to recognize these words subsequently. If this task is given in the context of a German course, this external condition may trigger the interpretation of stimulated recall (What does “task” mean in German? – “Aufgabe”). In other contexts different interpretations may be evoked. Thus, internal and external conditions are assumed to fill the missing or ambiguous slots of the COPES script for a given task. Therefore, task definitions are inherently idiosyncratic as they are constructed by partly using

information stored in the individuals' long-term memory (Winne, Jamieson-Noel, & Muis, 2002; Winne, 2006).

#### *Goal Setting and Planning – The Second Stage of Studying*

Based on the task definition generated in stage one, learners select or generate idiosyncratic goals and construct an elaborate plan for addressing the study task during the stage of *goal setting and planning* of the COPEs-model (Winne & Hadwin, 1998, Table 1).

Regarding *goals*, a multivariate profile of goals is generated which serves as monitoring standard in the subsequent learning stages (see paragraph about standards). Learners can adopt different kinds of goals simultaneously (Butler & Winne, 1995; Winne, 1995). Therefore, learners need to metacognitively regulate their behavior with regard to conflicting goals and sequencing goals. Goals can be very idiosyncratic because learners are assumed to “always have latitude to select goals, both within the confines of an assigned task and orthogonally to that task” (Butler & Winne, 1995, p. 256; also Anderson et al., 2001; Winne, 1997). More specifically, learners' goals might be impacted by *external conditions* as well as by learner characteristics (*internal conditions*). To give an example for the impact of internal conditions: A teacher might have set a task, such as reading a chapter, with the goal of learning. Although a disinterested student (internal condition: interest) might recognize the teacher's goal, she may consciously decide just to skim the text without paying real attention because she did not adopt the teacher's goal. Goals are also evaluated with regard to effort-utility payoffs in this stage: If the task is judged to be too complex, the effort to reach the “given” goal judged too high, the incentive of the goal judged too low, or the ability to reach the goal judged too low, the learner may abandon the whole task.

Regarding *plans*, once idiosyncratic goals for learning are selected, learners may either automatically retrieve tactics or strategies that are coupled with that kind of goal due to their expertise with such kinds of tasks or they may have to engage in effortful decision making by weighing the utilities of different options (Winne & Perry, 2000; Winne, Jamieson-Noel, & Muis, 2002). Winne (1997) argues that each time the learner considers a specific tactic the learner's memory system will automatically create an outcome expectation which activates further automatic memories (AEIOU: *attributions, efficacy expectations, incentives, outcome expectations, and utility*; Winne, 2001; Winne & Marx, 1989). Consequently, each tactic considered for task completion is assumed to be accompanied by such a bundle of “hot” information (Zajonc, 1980) which augments “cold” rational considerations (Anderson, 1991). Resulting, each considered tactic or strategy is associated with an overall perceived utility which in turn is assumed to lead to the selection of the tactic with the highest utility in the planning stage.

*Enactment – The Third Stage of Studying*

In the main *enactment* stage of the COPES-model (Winne & Hadwin, 1998, Table 1), the previously created plan of studying is carried out. The learner may enact all kinds of tactics or strategies to reach her goals (i.e., enact SMART operations, see paragraph about operations). During application of these operations learners are assumed to *metacognitively monitor* and *control* their learning process (see chapter about metacognitive processes). This stage of studying normally generates multiple products (see paragraph about products), for example new subject matter knowledge (internal) and notes or underlined text (external).

*Adaptation – The Fourth Stage of Studying*

The *adaptation* stage of the COPES-model (Winne & Hadwin, 1998, see Table 1) incorporates adaptations that exceed fine-tuning within one stage of studying. Thus, this stage has a backward and forward reaching nature that affects learners' more permanent approaches to studying. Two kinds of adaptation are described: One kind results in a large scale adjustment in how a learner coordinates studying activities across *multiple stages*. The other kind of adaptation affects *future* tasks by making changes to the cognitive structure based on the current studying experience (e.g., changes in beliefs, dispositions or motivational orientation; Winne & Perry, 2000). For example, learners may permanently change their knowledge about a specific learning tactic if it failed in one task. To come back to the example of the "newly learned definition": A student who regularly uses repetition as her default tactic might be introduced to the strategy of imagery. Thus, she might realize the low utility of repetition and might deliberately abandon the repetition strategy from her repertoire by assigning a very low utility to that strategy. Based on these large scale adaptations future studying tasks will be addressed. Therefore, these adaptations constitute the mechanism by which learning styles and dispositions are built up. This adaptation stage of studying is optional, not all learners will make adaptations in all studying experiences (Winne & Perry, 2000; Winne, 2001; Winne, Jamieson-Noel, & Muis, 2002).

Table 1: Facets of the COPEs typology (Table adapted from Winne & Hadwin, 1998, p. 282)

Facet	Definition	Stage 1: Task Definition	Stage 2: Goals & Plan	Stage 3: Enactment	Stage 4: Adaptation
<b>Conditions</b>	conditions that affect how the task will be engaged, including conditional knowledge (IFs in IF-THEN rules)	e.g., time constraints; available resources; knowledge of tactics; task knowledge; subject matter expertise	Stage 1 conditions + Stage 1 products	Stage 1 & 2 conditions + Stage 2 products	Stage 1-3 conditions + Stage 1-3 products
<b>Operations</b>	cognitive processes, tactics, and strategies the student engages to address the task	SMART operations: Searching; Monitoring (including tactics such as self-questioning, charting similarities and differences); Assembling (including tactics such as elaborating, integrating); Rehearsing (including tactics such as copying notes, re-reading); Translating (including tactics such as making diagrams, 1st-letter mnemonics)			
<b>Products</b>	information created by operations	Perception of what the task is, it's COPEs	standards for enacting tactics (e.g., speed, adapted to internal conditions) plan for coordinating study tactics	traces of study tactics (e.g., margin notes) new/re-organized subject matter knowledge	Updates to COPEs that distinguish tasks
<b>Evaluations</b>	feedback about products, either generated internally by the student or provided by external source(s)	Judgments about: - task understanding - COPEs of the task  Comparisons with other sources of information	Judgments about: - complexity/difficulty - incentive of goal - ability to carry out plan - effort required	Judgments about: - learning (calibration) - utility of tactics - efficacy - attributions	Judgments about: - "distance" between prior version and adaptation - predicted effects of adaptation Other evaluations must await future tasks
<b>Standards</b>	criteria against which products are monitored	grading criteria past performance	Stage 1 standards and products effort/utility thresholds motivation orientation	Stage 2 standards and products + object-level standards & meta-level standards	Stage 1-3 standards and products

### 2.1.3 Metacognitive Processes in the COPES-Model

Superlative descriptions underline the crucial importance of metacognitive processes in the COPES-model (Winne & Hadwin, 1998): *Metacognitive monitoring* is labeled “critical gateway” for self-regulated learning (Winne & Perry, 2000) or the “pivot on which self-regulated learning turns” (Winne, 2001). Furthermore, it is also noticeable because metacognitions permeate all previously described constituents of the COPES-model: Most facets for learning are either inputs or outputs of metacognitive processes and metacognitive processes play an important role in all stages of studying (see previous chapters). The notion of metacognition can be mainly traced back to Flavell (1976, 1987, 1992). Metacognitions are usually defined as “cognitions about one’s own cognitions”. Thus, they are not qualitatively different from other cognitions, only relationally different as they have the individuals’ own cognitions as an object (Nelson, 1999; Nelson & Narens, 1994). The dimensions within the construct of metacognition are discussed controversially (Efklides & Vauras, 1999; Brown, 1987; Boekaert, 1999). However, most often metacognitive *knowledge* (e.g., about learning strategies) can be distinguished from metacognitive *skills* (e.g., planning). Within the COPES-model two kinds of metacognitive processes are specified, metacognitive *monitoring* and metacognitive *control*. Both can be considered metacognitive skills (Schraw, 1998; Desoete, Roeyers, & Buysse, 2001; Veenman & Elshout, 1999).

The conceptualization of metacognitive monitoring and controlling in the COPES-model is similar to a model of metacognition by Nelson and Narens (1994). Within this model, object-level information refers to knowledge or cognitive operations, whereas meta-level information is assumed to have cognition as its object (e.g., metaknowledge about the utility of cognitive operations). The processes of metacognitive monitoring and controlling are operationalized in terms of “information flow”: During metacognitive monitoring information from the object-level updates meta-level information, thus information flows bottom-up. During metacognitive control the reverse process happens: Based on meta-level information top-down activation is initiated, for example cognitive strategies that are executed subsequently. This kind of model is also often referred to as “discrepancy-reduction-model” (Dunlosky & Hertzog, 1997; Dunlosky & Thiede, 1998; Otero, 2002; Puustinen & Pulkkinen, 2001; Thiede & Anderson, 2003) because metacognitive control is assumed to be mostly exerted to reduce discrepancies that are detected due to metacognitive monitoring. This notion is congruent with related theories such as Carver and Scheier’s (1990) model of regulated behavior or the conceptualization of behavior as TOTE (test-operate-test-exit) cycles (Miller, Galanter, & Pribram, 1960).



### *Metacognitive Monitoring*

Metacognitive *monitoring* is usually defined as “one’s on-line awareness of comprehension and task performance” (Schraw, 1998, p. 115). Most researchers assume that the evaluative monitoring process constitutes a comparison between an internal standard and a product of learning (Otero, 2002; Thiede & Anderson, 2003). Thus, metacognitive monitoring helps to detect potential problems (Baker, 1985; Garcia-Arista, Campanario, & Otero, 1996; Dunlosky, Rawson, & Hacker, 2002; Hacker, 1997, 1998; Kinnunen & Vauras, 1995; Otero, 1998, 2002; Wagoner, 1983; Zimmerman, 1990). The definition given in the COPES-model is consistent with these notions: Winne and Hadwin (1998) consider monitoring to be a cognitive operation that has two inputs and creates one output. One input is an object-level multivariate profile of attributes for a product (see paragraph about products). The second input is a meta-level multivariate profile of attributes that constitutes the internal standard for that particular product (see paragraph about standards). As a result of metacognitive monitoring a third multivariate profile of attributes is created that records discrepancies between the object-level information and the meta-level information: a profile of evaluations (see paragraph about evaluations).

### *Metacognitive Control*

Metacognitive *control* on the other hand refers to all activities enacted to actively regulate one’s own behavior, for example to compensate for problems detected by metacognitive monitoring (Baker, 1985; Dunlosky, Rawson, & Hacker, 2002; Hacker, 1997, 1998; Kinnunen & Vauras, 1995; Otero, 2002). The view of the COPES-model is consistent with this notion: According to Winne and Hadwin (1998), the discrepancies detected by metacognitive monitoring can be modeled as IFs in IF-THEN rules, while metacognitive control can be modeled as THENs. More specifically, IF a discrepancy is detected – either by internal or external feedback –, THEN metacognitive control is exerted. Thus, metacognitive control may be exerted to fine-tune the enacted tactics and strategies, for example “to make mid-course adaptations” (Winne, 1996, p. 331). Metacognitive control changes the task space at the object-level. In the COPES model, two forms of metacognitive control are assumed: toggling and editing. With *toggling* a learner turns cognitive operations on or off, for example a learner stops using the ineffective tactic of repetition for memorization and instead starts to use the more effective tactic of imagery. *Editing* on the other hand implies changes of operations, which can be achieved via (1) accreting (generalizing or specializing an operation), (2) tuning (changing sequence of operations) or (3) restructuring (inventing new tactics). In the above mentioned example of a “newly learned definition” the learner could, for example, decide to specialize her use of the repetition tactic by only using it if she is unable to create a meaningful image for imagery.

### 2.1.4 Conclusion

The COPES-model is highly popular in education and educational psychology and fairly well grounded in empirical research. The most comprehensive description of the COPES-model (Winne & Hadwin, 1998) was cited twenty-four times, but all accounts of the COPES-model together elicited a total of 249 unique citations (counted September 2006 in the Science Citation Index (SCI)). With regard to an empirical basis a recent review of the COPES-model (Greene & Azevedo, 2007) demonstrates that most of its assumptions are consistent with empirical findings. For example, empirical studies support the notion of learners' idiosyncratic task definitions (Cordón & Day, 1996; Van de Watering & Van der Rijt, 2006), they support the impact of internal and external conditions (Ainley, 1993; Erlbaum, Berg, & Dodd, 1993; Wolters, Yu, & Pintrich, 1996), and they validate the assumption of an adaptation stage (Eisenberger, Masterson, & McDermitt, 1992; Rabinowitz, Freeman, & Cohen, 1992). To give a more specific example: Van Meter, Yokio, and Pressley (1994) showed that students' self-reported note taking in the enactment stage varied depending on the lecturers' delivery speed and organization (external conditions) and their own domain knowledge (internal condition).

Because of these benefits of the COPES-model it was selected as a general framework for this thesis. However, the COPES-model is underspecified with regard to some open issues outlined below. These open issues will be explicitly addressed in this thesis.

#### *(How Exactly) Do Learners Adapt their Learning to External Conditions?*

Adaptivity to *external conditions* is conceptualized as a core element of self-regulated learning within the COPES-model. Learners are assumed to monitor external conditions and exert metacognitive control to adequately adapt their learning processes. Additionally, better adaptation is supposed to be beneficial for learning outcome. This notion is endorsed by other theoretical models of self-regulated learning (Boekaerts & Niemivirta, 2000; Borkowski, Chan, & Muthukrishna, 2000; Pintrich, 2000, 2002, 2004; Pintrich, Wolters, & Baxter, 2000; Winne & Hadwin, 1998; Zimmerman, 2002; for an overview see Puustinen & Pulkkinen, 2001), metacognition (Bannert, 2003; Nelson, 1999; Nelson & Narens, 1994), test preparation (Broekkamp & van Hout-Wolters, 2006) or reading (Reynolds, 1992). Within this thesis, learners' adaptivity to *task complexity* (chapter 2.3) and *text complexity* (chapter 2.4) will be scrutinized. These external conditions were selected based on a conceptual analysis of learning with hypertexts: For example, task complexity might determine the overall learning strategies while text complexity might determine which kind of reading strategies are enacted for specific hypertext pages.

Note that these adaptation processes are described quite abstractly in the COPES-model. It does not detail a functional model of *how exactly* these adaptation processes are executed. Therefore, it is only suited as a general research framework. A more detailed conceptualiza-

tion, an adequate terminology as well as a suited research methodology can be transferred from the *traditional calibration paradigm*, at least if the traditional definition of the construct “calibration” is extended (chapter 2.2). With this methodology it is possible to diagnose if learners’ self-regulated learning process is in fact systematically related to the selected external conditions. Consequently, adaptation processes with regard to external conditions will be re-conceptualized as calibration processes. Accordingly, the *conditions and processes conditions of metacognitive calibration* with regard to task complexity and text complexity constitute one core research questions of this thesis.

*(How Exactly) Do Internal Conditions Impact these Adaptation Processes?*

Additionally, the COPES-model acknowledges the impact of *internal conditions*. This focus is also endorsed by other theoretical models of self-regulation and metacognition (Broekkamp & van Hout-Wolters, 2006; Butler & Cartier, 2004; Reynolds, 1992). Within this thesis two internal conditions will be scrutinized: *prior domain knowledge* (chapter 2.5) and *epistemological beliefs* (chapter 2.6). These internal conditions were selected because the COPES-model “acknowledges the pivotal role of prior knowledge, [...] and] epistemological beliefs, [...] across all phases of studying” (Hadwin & Winne, 2001, p. 324).

Note that the impact of these learner characteristics is also described quite abstractly in the COPES-model. It does not detail a functional model of *how exactly* this assumed impact works. Therefore, diagnosing the complex interactions between external conditions and internal conditions is a second main focus of this thesis. Consequently, this thesis will determine if internal conditions have a *main* effect independent of external conditions or if internal and external conditions significantly *interact* (e.g., prior domain knowledge might be more influential for more complex tasks). To summarize, these *conditions and processes of metacognitive adaptation* will be scrutinized as another core research questions of this thesis.

*Do Conditions Impact All Stages Alike?*

On the one hand, the COPES-model assumes similar cognitive processes in all stages of studying (e.g., the same COPES facets); on the other hand it assumes distinct stages. However, it is unclear in which stages and how exactly internal and external conditions exert their impact. For example Greene and Azevedo (2007) wondered with regard to the impact of the external condition context “in what phase(s) these effects occur” (p. 344).

To detangle the impact of conditions in different stages of learning, the empirical studies of this thesis will either exclusively concentrate on the *preparatory* stages (study I) or primarily focus on the *enactment* stages (study II and study III). This subdivision into two broader stages is congruent with earlier versions of the COPES-model (Winne, 1996, 1997), but also compatible with Winne and Hadwin’s (1998) notions, at least if the stages of *task definition* and *goal setting and planning* are combined and the *adaptation* stage is considered optional. It makes sense that processes necessary for the preparatory stages of learning (e.g., knowl-

edge about tasks or about the utility of tactics) might be different from those needed in the enactment stage (e.g., restrained by working memory capacity).

### *Is More Flexible and Accurate Adaptation Beneficial for Learning?*

The COPES-model posits that more flexible and accurate adaptation to external conditions (e.g., to *task complexity* and *text complexity*) should be beneficial for learning. However, given that the COPES-model can be considered an idealized and prescriptive account rather than a realistic description of learning, this is an open issue.

For example, empirical results demonstrate that learners do not always select the “objectively best” tactics (Kuebli & Fivush, 1994), probably because they have insufficient knowledge about the utility of tactics. Neither all students nor all instructors know about the spacing effect, that is that learning can be improved by distributing study sessions in comparison with massed studying (Dempster, 1988), or that an integrated summary over multiple sources is more effective than multiple single summaries (Spurlin, Danserau, O’Donnell, & Brooks, 1988). Additionally, empirical results also indicate learners’ poor spontaneous metacognitive monitoring (Anderson & Beal, 1995; Baker, 1985; Hacker, 1997; Hare & Bouchard, 1985; Markman & Gorin, 1981; Winograd & Johnson, 1982; Schraw, 1998). However, metacognitive monitoring was consistently positively associated with enhanced understanding (Kinnunen & Vauras, 1995), better writing (Beal, 1996; Hacker, 1997), better problem solving (Davidson & Sternberg, 1998; Metcalfe, 1992) and better academic achievement in general (Otero & Campanario, 1992). Furthermore, monitoring ability increases with age and ability (Baker, 1984, 1985; Garner, 1980; Hacker, 1997) and can be trained (Schraw, 1998; Thiede & Anderson, 2003).

Based on these findings, it is feasible that the quality of learners’ calibration to external condition (chapter 2.2) might have no relation to the learning outcome (e.g., if learners do not systematically adapt their learning to external conditions), might be beneficial (e.g., if the assumptions of the COPES-model are true), or might even be detrimental. Thus, the empirical studies within this thesis will additionally scrutinize *potential determinants of the learning outcome*, especially if more flexible and accurate adaptation to external demands is beneficial for learning (“*To calibrate or not to calibrate?*”).

## **2.2 Metacognitive Calibration**

Traditionally, the calibration paradigm was used to investigate the accuracy of learners’ metacognitive *monitoring*. However, in this chapter I will argue that this traditional conceptualization of calibration is flawed – at least if the whole self-regulated learning process is focused. Thus, two extensions of the traditional construct “calibration” will be suggested in order to alleviate these problems and to enable the investigation of the core questions of

this thesis within this paradigm. A further advantage of these extensions is the potential transfer of methodology to a new application context: The methodology of the traditional calibration paradigm is well-suited to analyze the processes and conditions of metacognitive adaptation (see previous chapter) in detail.

### 2.2.1 Selective and Critical Review of Traditional Research on Calibration

To give a short overview of calibration in its traditional sense: *Calibration* is most often defined as congruence, alignment or match between students' *subjective judgments* with regard to learning, memory or performance and their performance on a *criterion test* (Dunlosky & Hertzog, 2000; Garavalia & Gredler, 2003; Glenberg & Epstein, 1995, 1987; Lin & Zabrocky, 1998; Lodewyk & Winne, 2005; Nelson, 1996; Nelson & Dunlosky, 1991; Nietfeld, Enders, & Schraw, 2006; Schraw, 1995; Schraw & Roedel, 1994; Stone, 2000; Weingardt, Leonesio, & Loftus, 1994; Wiley, Griffin, & Thiede, 2005; Winne & Muis, 2004). Thus, it characterizes how aware individuals are of their own internal processes, for example of what they do and do not know or what they do and do not do (Glenberg & Epstein, 1985; Stone, 2000). Such accurate monitoring is assumed to be a necessary precondition for successful learning (Lin & Zabrocky, 1998; Stone, 2000; Wiley, Griffin, & Thiede, 2005; Winne, 1995, 1996). To illustrate this traditional calibration paradigm, an example of a typical study will be given (Nelson & Dunlosky, 1991): Participants memorize pairs of unrelated nouns such as "ocean – tree" (paired associates), subsequently participants give subjective confidence ratings for recall of each paired associate ("How confident are you that in about ten minutes from now you will be able to recall the second word of the item when prompted with the first?"), for example "ocean - ?") and finally their objective recall is measured. In this example, calibration is defined as the match between participants' confidence judgments (subjective metacognitive judgment) and their performance on the recall test (criterion test). Methodologically, calibration can be captured either by the number correct judgments (*absolute* calibration) or by within-subject measures of association between subjective judgments and performance on a criterion test (*relative* calibration: e.g., correlation). Throughout the research literature, a wide range of types of subjective judgments and objective criterion tests have been employed.

Common subjective metacognitive judgments include (but are not limited to): (1) predictions of recall for separate paired associates (see example above; Nelson & Dunlosky, 1991; Dunlosky & Hertzog, 2000), (2) prediction of correctly answering separate inference questions (after reading short texts participants have to give confidence ratings that they will be able to judge the correctness of an inference; Glenberg & Epstein, 1987), (3) prediction of learning result (participants are asked to predict their percentage of correct answers; Burson, Klayman, & Larrik, 2006; Dunlosky & Hertzog, 2000); (4) prediction of own percen-

tile rank (participants are asked to estimate the percentile rank in relation to their peers; Burson, Klayman, & Larrik, 2006); (5) postdiction of achievement for separate achievement tests or items (after performance on the criterion test participants rate their confidence in performance for each item; Dunlosky & Hertzog, 2000; Glenberg & Epstein, 1987; Nietfeld & Schraw, 2002; Pressley & Ghatala, 1988; Schraw & Roedel, 1994; Weingardt, Leonesio, & Loftus, 1994; Winne & Jamieson-Noel, 2002; Winne & Muis, 2004), and (6) postdiction of achievement across multiple achievement tests (after performance on the criterion test participants are asked to postdict the percentage correct answers; Burson, Klayman, & Larrik, 2006; Dunlosky & Hertzog, 2000; Fitzgerald, White, & Gruppen, 2003; Schraw & Roedel, 1994). It can be inferred that subjective metacognitive judgments vary on multiple dimensions, two of the most important are time of judgment, either prediction (see examples (1) - (4)) or postdictions (see examples (5) - (6)), and granularity of judgment, either specific to one item (see examples (1), (2) and (5)) or more global judgments across multiple items (see examples (3), (4), and (6)).

Common criterion tests include (but are not limited to): (1) cued recall of paired associates (after memorizing paired associates participants have to complete a cued recall test; Dunlosky & Hertzog, 2000; Nelson & Dunlosky, 1991), (2) verbatim or idea recognition (after reading a text participants have to decide if presented sentences or ideas were present in the text; Glenberg, Sanocki, Epstein, & Morris, 1987; Weingardt, Leonesio, & Loftus, 1994), (3) quizzes (without prior learning participants have to answer quiz questions about topics such as university trivia; Burson, Klayman, & Larrik, 2006), (4) open questions that are scored as either correct or incorrect (without prior learning participants have to answer open questions about topics such as general knowledge, word knowledge, or mathematic knowledge (“What number results if you multiply a number by its inverse?”); Winne & Muis, 2004), (5) multiple-choice questions (without prior learning participants have to answer multiple choice questions about topics such as opposites, reading comprehension, spatial judgments, probabilities, or analogies (“The relationship of intruder – privacy is analogue to a) blot – ink, b) ripple – calm, c) noise – clamor, d) animal – forest, or e) hermit – solitude?”); Nietfeld & Schraw, 2002; Pressley & Ghatala, 1988), Schraw & Roedel, 1994), (6) inference verification tasks (after reading a text participants had to judge if an inference drawn from the text was correct or incorrect; Glenberg & Epstein, 1987) or (7) tests of deeper understanding of the learned material (for example, Fitzgerald, White, and Gruppen (2003) used practical medical exams as criterion tests). Summarizing, these criterion tests differ foremost in the depth of understanding that is required to answer the criterion test. While simpler criterion tests require only recall or recognition (e.g., examples (1) and (2)), other criterion tests require deeper understanding such as comprehension (example (5)), independent inferences (example (6)) or application of the acquired knowledge (example (7)). Furthermore, some criterion test measure newly learned content (e.g., examples (1), (2), (6), (7)), while other measure prior knowledge (e.g., examples (3), (4), (5)).

### *Frame of Reference Problem*

The differences within the criterion tasks in depth of required understanding potentially lead to a problem with regard to the corresponding subjective metacognitive judgments that can be discussed under the label *frame of reference problem*. This problem reflects the fact that participants' task understanding may not always reflect objective task demands but can be very idiosyncratic, especially for more ill-defined and complex tasks (see description of the *task definition* stage in the COPEs-model, chapter 2.1.2). Therefore, participants might have different criterion tasks in mind when making their metacognitive judgments. To give an example: If a participant thinks she will be tested on verbatim recall and gives her judgments according to this frame of reference, but later on is presented with a criterion test that affords inferences and transfer, then low calibration is not surprising. Empirical findings support the *existence* of this frame of reference problem on several levels.

First, this argument is underlined by empirical results within the traditional calibration paradigm showing that, in general, *postdictions* are more accurate than *predictions*. In a typical prediction task, learners make their metacognitive judgments after learning, but before taking the objective criterion test (Nelson & Dunlosky, 1991). Postdictions on the other hand are given after both learning as well as the criterion tasks have been completed. For example, in the study of Winne and Jamieson-Noel (2002) sixty-two students used the software PrepMate to study a text chapter on lighting. Afterwards they completed an achievement test that contained six criterion tasks. *After* each task, students gave subjective judgments ("This question was worth 5 (or 10) points. Based on your answer, what would you give yourself?"). Median calibration, operationalized as within-subject correlation between achievement on the criterion test and postdicted achievement estimates, was quite substantial with  $r = .88$ . To underline this argument, Maki (1998) reports more aggregate data: 25 studies of calibration were conducted in her lab. For predictions a mean Gamma correlation of  $G = .27$  was found. The corresponding correlation for postdictions was considerably higher with  $G = .48$ . Therefore, Maki (1998) argues that in postdiction tasks the exact nature of the tasks is known before the participants are required to make their judgments while in prediction tasks participants also have to predict the nature of the task which potentially adds more noise to participants' predictions of performance. This argument is further supported by studies that show that prediction accuracy can be improved if practice questions of the same nature as the criterion tasks are given to participants (Maki & Serra, 1992; Wiley, Griffin, & Thiede, 2005).

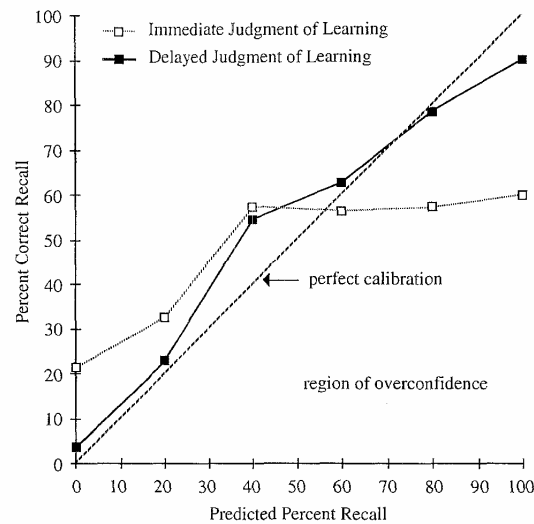


Figure 2.2-1: *Calibration* between JOLs (judgments-of-learning, predicted percent recall) and performance on a criterion test (percent correct recall) for immediate and delayed JOLs (Figure taken from Nelson & Dunlosky, 1991, p. 269).

Second, this argument is underlined by empirical results within the traditional calibration paradigm showing that calibration also varies depending on the *complexity of the criterion task*: Judgments with regard to paired associates (simple, well-defined) are generally more accurate than judgments with regard to learning from written text (more complex, ill-defined; Wiley, Griffin, & Thiede, 2005). First, consider research concerned with *predictions*: In *paired associate learning* near perfect calibration can be achieved under some conditions. For example, Nelson and Dunlosky (1991) demonstrated that calibration accuracy can be improved if predictive confidence ratings are not given immediately after memorizing an item ( $G = .38$ ), but after a delay ( $G = .90$ ; Figure 2.2-1). Other findings show that accuracy of judgments in paired associate learning tasks can also be improved by practicing and active construction (Wiley, Griffin, & Thiede, 2005). In *learning from written text tasks* on the other hand perfect calibration is harder to achieve (Lin & Zabricky, 1998). In a typical study Glenberg and Epstein (1985, experiment 1) asked participants to read 15 expository texts. After reading each text, students predicted their ability to answer an inference question about a central point within the text. These confidence ratings were neither predictive if given immediately following each text ( $r = .07$ ) nor if delayed until after all texts had been read ( $r = .04$ ). Later reviews (Wiley, Griffin, & Thiede, 2005) argued that these inference questions varied in difficulty and thus were hard to predict by the readers: Even if a student correctly predicted the nature of one inference task that interpretation might be quite wrong for the next task. Further empirical evidence shows that even *postdictions* for written texts seem to be less accurate than for simpler, well-defined tasks. Pressley and Ghatala



(1988), for example, confronted their participants with three kinds of multiple-choice questions: opposites, analogies, and reading comprehension questions. Subsequently, they asked their participants to postdict their achievement. Results indicate significant differences in postdiction: Judgments for opposites ( $G = .70$ ,  $G = .78$  for two parallel tests forms) and analogies ( $G = .71$ ,  $G = .70$ ) were significantly more accurate than those for reading comprehension ( $G = .31$ ,  $G = .46$ ) which was the most complex and ill-defined tasks.

Third, this argument is underlined by empirical results within the *error detection paradigm* which show that learners significantly differ in their perception of the complex task “text comprehension” (Anderson & Beal, 1995; Baker, 1984a, 1984b, 1985a, 1985b; Beal, 1996; Garcia-Arista, Campanario, & Otero, 1996; Garner, 1980; Hacker, 1997; Hare & Borchart, 1985; Otero & Campanario, 1992; Plumb, Butterfield, Hacker, & Dunlosky, 1994; Walczyk & Hall, 1989; Winograd & Johnston, 1982; Zabucky & Moore, 1989). Within the error detection paradigm errors are deliberately inserted into a text. Participants read the text and are prompted to report their difficulties. Successful detection of inserted errors is interpreted as a sign of good metacognitive *monitoring of comprehension*, while a failure to detect errors is assumed to be caused by a monitoring deficit. Furthermore, to explore participants’ underlying *standards* for comprehension, different kinds of errors can be inserted in the text material. The following example will illustrate a typical study: Zabucky and Moore (1989) classified 149 fourth graders, 120 fifth graders and 125 sixth graders as poor, average or good readers. Each student read 8 expository texts, each 6 sentences long. Four versions of each text were created by systematically varying only the fourth sentence: (1) intact version (no problem), (2) inconsistent version (contradiction with 2<sup>nd</sup> sentence), (3) falsehood version (prior knowledge violation) and (4) nonsense word version (contained one nonsense word). Students received either a general instruction to underline all problems or a specific instruction to underline nonsense words, inconsistencies or falsehoods. Results show main effects of all variables. Higher grades, better reading ability and a specific instruction lead to better error detection. Furthermore, nonsense words and falsehoods were significantly more frequently detected than inconsistencies. In this study as well as in other studies within the error detection paradigm good students employed a *diversity* of standards for comprehension monitoring and were also more likely to employ standards of deeper *elaboration*, for example semantic standards (Baker, 1984a, 1984b, 1985a; Hacker, 1997). Poor students on the other hand used *fewer* standards and primarily more *superficial* lexical or syntactical standards (Baker, 1984a, 1984b; Hacker, 1997). Explicit prompting to look out for special kinds of inserted errors raised error detection rates, but did not change the overall pattern of results (Anderson & Beal, 1995; Baker, 1984a). This can clearly be interpreted as an indicator of different interpretation of the task “comprehension”. If these findings are transferred to the traditional calibration paradigm, it is likely that participants in such studies might work in different *frames of reference* for a specific task.

Fourth, this argument is underlined by empirical results that show directly that different people (e.g., teachers and their students) differ in their *task perception* as well as in their *perception of learning material* (Broekkamp, van den Bergh, van Hout-Wolters, & Rijlaarsdam, 2002; Broekkamp, van Hout-Wolters, Rijlaarsdam, & van den Bergh, 2002; Broekkamp, van Hout-Wolters, van den Bergh, & Rijlaarsdam, 2004). For example, Van de Watering and van der Rijt (2006) had groups of teachers (group size:  $n = 4 - 7$ ) construct batteries of multiple-choice questions. Each teacher rated each question on a 3-point scale (1 = difficult - 3 = easy). These batteries were administered to students who also rated question difficulty on the same scale. Accuracy of teachers' and students' task perception was computed by subtracting individual estimates from the overall difficulty indices of the items (proportion of correct answers). Teachers considered items easier than indicated by the difficulty index (differences ranged from .05 to .06). Students on the other hand considered items – especially easy ones – more difficult than indicated by the difficulty index (differences ranged from -.11 to -.17). These as well as other empirical results lead to the conclusion that “learners do not approach activities as teachers intend” (Butler & Cartier, 2004, p. 1731). If these findings are transferred to the traditional calibration paradigm, it is likely that participants in such studies work in different *frames of reference* from the researchers.

#### *Questionable Construct Validity of Metacognitive Judgments*

The construct validity of metacognitive judgments used in the calibration paradigm can be questioned because metacognitive processes might be automated, *unconscious* and they might not naturally occur on such *fine-grained* level (ecological validity). Consequently, it can be concluded that metacognitive judgments may not constitute the best-suited proxy for *ongoing metacognitive monitoring* processes during self-regulated learning but might rather reflect students' *metacognitive knowledge*.

Theoretically, this notion of potentially unconscious metacognition is supported by arguments within the COPES-model (Winne & Hadwin, 1998; see previous chapter). In contrast to common conceptualizations of metacognition (Carver & Scheier, 1990; Garner, 1980; Hacker, 1998; Hasselhorn, 1998; Nelson, 1999; Zimmerman, 1989), the COPES-model explicitly considers the possibility that some metacognitive processes may be executed without deliberate and conscious control (also Schunn, Lovett, & Reder, 2001; Sternberg, 1998; Veenman, 2003). Based on empirical research that showed that learners can metacognitively regulate their behavior without conscious deliberation (Kanfer & Ackerman, 1989), the COPES-model assumes that metacognitive processes are subject to automation by practice. More specifically, Winne (1995, 1996, 1997) proposes a three-step automation process based on Fitts (1964): (1) First, learners only possess declarative conditional and action knowledge in an IF-THEN format (cognitive stage), (2) by practicing these two kinds of knowledge then become linked and form propositions (associative stage) that (3) in the end become automated by repeated use (automated stage). This view

also implies that as long as metacognitive processes are conscious and deliberate they require some effort and cognitive resources and compete with the primary cognitive processes for working memory resources.

Empirically, this notion is further corroborated by studies that show that learners might not have access to their own metacognitive processes: Learners' self-reports about their metacognitions and learning strategies often show little relation to their real learning processes. The construct validity of self-reports is low in comparison with concurrent assessments of learning (Artelt, 2000; Bannert, 2003; Veenman, 2003; Winne & Jamieson-Noel, 2002): Learners sometimes do not do what they say they do. To give an example: Artelt (2000) compared students' ( $n = 250$ ; grade 4, 6, and 8) retrospective self-reports about strategic learning and their actual use of learning strategies during text studying. No relations could be found between retrospective self-reports and the actual use of strategies.

As a consequence of such automated metacognitive processes, forcing students to generate metacognitive judgments could lead to artificial inferences. To borrow an explanation from the think aloud paradigm about so-called level-3-verbalization (Ericsson & Simon, 1980): Metacognitive judgments are not necessarily contained in short term memory during learning due to high automation. Thus, students may have to actively re-construct this information and this process might be biased and artificial. However, considering which sources learners might utilize to re-construct those judgments, these judgments are most likely based on their metacognitive knowledge.

#### *Insufficient Ecological Validity*

Two concerns with regard to the ecological validity of calibration can be raised within the traditional calibration paradigm: First, the traditional calibration paradigm captures students' calibration to *internal* standards (students' *personal* judgments are compared with their *personal* performance) but does not capture students' calibration to *external* demands. However, a student might be internally well-calibrated (e.g., be totally aware of which items she will recall) but still be ill-calibrated with regard to external demands (e.g., totally unaware of which items will be important on an upcoming test). Thus, perfectly calibrated students in the sense of the traditional calibration paradigm might still be unsuccessful self-regulated learners and consequently might also perform poorly on tests.

Second, calibration in the traditional sense only captures a limited fraction of the whole self-regulated learning process: *metacognitive monitoring*. Learners only have to master two of four necessary affordances of successful self-regulated learning in order to achieve accurate calibration (Butler & Winne, 1995; Winne, 2005, 2006; Winne & Jamieson-Noel, 2003): Learners must (1) recognize presented cues (*monitor* cues that indicate future recall, for example close semantic association between two words like "cat – dog"), (2) and must interpret these cues correctly (*metacognitively know* under which conditions recall is most likely, for example for closely associated words). However, learners do not have to (3) actively enact

or *produce* operations and therefore they also do not have to (4) be *motivated* enough to spent the necessary effort. Even though this kind of monitoring captured by the traditional calibration paradigm may be a necessary precondition for successful learning, it does not guarantee success (Lin & Zabucky, 1998). For example, a student might be well-calibrated (e.g., be totally aware of which items she will recall) but still unable to *enact* adequate remedial control strategies. Thus, crucial parts of the self-regulated learning process can not be investigated with this paradigm (e.g., *metacognitive control strategies*).

### 2.2.2 An Extended Definition of the Construct Calibration

Based on the previously presented critical review of the traditional calibration paradigm and based on the affordances of the core research questions of this thesis, two extension of the construct definition of calibration will be suggested: The construct of calibration should be extended to also encompass the question of students' (1) *active adaptation* to (2) *external task demands*. If calibration in this extended sense is investigated most outlined problems associated with the traditional calibration paradigm can be alleviated.

#### *Calibration to External Conditions – Proposed Extension of the Criterion*

As detailed in the critical paragraph on ecological validity, ultimately educators as well as researchers are not interested in students who are absolutely aware of their *own* cognitive processes or performance, but instead learners also – or probably primarily – should be aware of *external demands* in order to be successful self-regulated learners. Thus within this proposed extended construct definition, calibration should not only refer to a comparison between students' metacognitive judgments and *their own performance (internal criterion)*, but also to a comparison between students metacognitive judgments and *external criteria*.

For example, one important external characteristic that should be monitored by students is the learning task. But the demands of learning tasks might be perceived quite differently by different learners as discussed under the label *frame of reference* problem. It was demonstrated that learners not only have to estimate their performance but also the nature of the criterion task, especially for predictions and especially for more complex criterion tasks. These idiosyncratic task definitions constituted unwanted *error variance* within the traditional calibration paradigm. On the other hand, if the proposed extension of the construct calibration is considered, this error variance can be considered a *primary research question*: How well do students' idiosyncratic task definitions match the objective properties of the tasks? How accurately calibrated are students' task perceptions to external demands? This research question is especially relevant because in most natural learning scenarios learners do not only face simple and well-defined tasks (e.g., vocabulary learning), but more frequently, learners encounter ill-defined and complex tasks such as having to write complex docu-

ments outlining and justifying their own point of view (e.g., essays, papers, theses). Thus, the ability to accurately judge external task demands is crucial for learning in the real world and constitutes a valuable goal of education.

#### *Calibration of Metacognitive Control – Proposed Extension of Subjective Judgments*

The critical review of the traditional calibration paradigm not only highlighted problems with the criterion task but also problems with the corresponding metacognitive judgments, (construct validity and ecological validity). Ultimately educators as well as researchers are not interested in students who are absolutely *aware* of their own cognitive processes but instead learners also – or probably primarily – should be able to *act* adequately upon their perceptions. Therefore, learners' enacted learning strategies should be captured as proxies of their metacognitive adaptation processes. Thus within this proposed extended construct definition, calibration should not only refer to a comparison between *students' metacognitive judgments* and external criteria, but also to a comparison between *students' enacted metacognitive control strategies* and external criteria.

Multiple arguments for this shift from monitoring to control can be given. First, as highlighted in the paragraph about the construct validity of metacognitive judgments, *ongoing metacognitive monitoring* processes might not be captured adequately by metacognitive judgments, possibly due to high automation. To exaggerate this argument: why should anyone be interested in accurate calibration between metacognitive judgments and criterion tasks, if metacognitive judgments do not validly represent the real metacognitive monitoring processes? The proposed shift to measuring active control strategies would be analogous to measuring students' concurrent learning activities (e.g., via think aloud protocols) instead of administering self-report questionnaires. In general this strategy is considered beneficial (Baker, 1985b; Veenman, 2003).

Second, as highlighted in the paragraph about the ecological validity of the traditional calibration paradigm, this research methodology is only able to give insight into a small fraction of the self-regulated learning process. The proposed extended construct definition of calibration on the other hand would enable researches to capture the *complete self-regulated learning process*. For example, all affordances of successful self-regulated learning could be considered (Butler & Winne, 1995; Lodewyk & Winne, 2005; Winne, 2005, 2006; Winne & Jamieson-Noel, 2003): Learners would not only have to (1) *monitor* presented cues (e.g., that indicate task complexity), (2) and possess *metacognitive knowledge* to interpret these cues correctly (e.g., memorization is not beneficial for a complex task). But learners would also have to be able to (3) actively apply or *produce* appropriate operations (e.g., strategies of deep elaboration) and also (4) be *motivated* enough to spent the necessary effort. In terms of the COPES-model (Winne & Hadwin, 1998) this extension of the construct “calibration” would allow to investigate all stages of the self-regulated learning process.

Additionally, this extension would constitute an example of the often proposed but rarely implemented *shift* in research from metacognitive *monitoring* to metacognitive *control* (Baker, 1985b; Son & Schwartz, 2002). Furthermore, the extension would also be in line with *everyday definitions* of the term calibration which not only stress the evaluative component but also the active control component (Wikipedia, the Free Encyclopedia: “adjusting the output [...] to agree with value of the applied standard”; Merriam Webster Online: “to adjust precisely”; WordNet Online: “fine-tune”). These definitions show that the concept of calibration usually not only denotes the measurement of deviation from an objective standard (see the traditional calibration paradigm), but also the active adjustment to an objective standard (see proposed extension of the construct calibration).

### 2.2.3 A New Research Strategy – Transfer of Methodology

In order to empirically investigate this extended notion of calibration an adequate methodology is needed. The traditional calibration paradigm offers a very specific methodology which constitutes a well-suited analytical tool to determine the core idea of calibration. An important advantage of these traditional measures is that they explicitly consider the *relation* between students’ metacognitions (traditional: metacognitive judgments; extended: metacognitive control) and a criterion (traditional: their own performance; extended: external criteria such as task complexity). Thus, this methodology should be retained to also investigate calibration within the proposed extended conceptualization. This traditional methodology will be reviewed and potential options for transfer will be discussed.

Within the traditional calibration paradigm, a distinction is often drawn between *absolute* accuracy and *relative* accuracy (Nelson, 1996; Nietfeld, Enders, & Schraw, 2006). Absolute accuracy focuses on a direct comparison between the absolute value of the metacognitive judgment and the absolute value of the criterion test. Therefore, it is often expressed as absolute degree of over- or underconfidence. Relative accuracy on the other hand is most often expressed by measures that capture the degree of association between metacognitive judgments and performance on the criterion tests across multiple items (e.g., correlations). In this thesis, the term *absolute calibration* will consistently be used to refer to absolute accuracy or to what Nelson (1996) termed calibration while the term *relative calibration* will be used to refer to relative accuracy or in Nelson’s (1996) words resolution. Sometimes, the related concept of *discrimination* is also considered as a necessary precondition for calibration in the traditional calibration paradigm. Furthermore, discrimination and calibration are often visualized with so-called *calibration graphs* within the traditional calibration paradigm.

### *Discrimination*

Within the traditional calibration paradigm, discrimination denotes students' ability to distinguish instances when a target event is going to occur from those when it is not (Weingardt, Leonesio, & Loftus, 1994). In eyewitness research for example, discrimination concerns the ability of a witness to distinguish situations where his or her judgments are more likely to be accurate from situations where they are less likely to be accurate (Weingardt, Leonesio, & Loftus, 1994). In this example, high discrimination would be diagnosed if the *metacognitive judgments* of a witness vary significantly. Thus, a statistical way to diagnose discrimination is given by t-tests (for dependent means) or (repeated-measure) ANOVAs. The performance on the *criterion test* is not taken into account to diagnose discrimination.

This methodology of discrimination can be transferred easily to investigate the extended notion of calibration. For example, if students' calibration to the external demand *task complexity* should be investigated, discrimination can be diagnosed if students derive significantly *different* task definitions and goals and plans for tasks of *different* complexity and if they enact significantly different strategies for those tasks.

To give an example of an empirical study using this research strategy (although the researchers never explicitly used the term discrimination): Hadwin, Winne, Stockley, Nesbit and Woszczyna (2001) showed that students discriminated between three different tasks. During a course in educational psychology students ( $n = 100$ ) filled in a self-report questionnaire three times: for reading for learning, for completing a brief essay, and for studying for an exam. The questionnaire asked for a frequency judgment about their application of tactics (e.g., structuring content) and resources (e.g., provided resources), and their selection of goals (e.g., selecting for depth). Subsequent ANOVAS showed main effects of context: "responses to self-report items about study tactics, selecting goals, and using external resources vary when study context varies" (Hadwin et al., 2001, p. 481). This study gives a first hint that university students can perceive external cues and recognize the different utility of tactics, resources and goals for different contexts. Students seem to be able to *discriminate* between tasks.

### *Absolute Calibration*

Within the traditional calibration paradigm different measures were suggested to capture students' absolute calibration. To evaluate the adequacy of these measures of calibration, Nelson (1984) suggested some desirable properties<sup>3</sup>: property # (1): as the ability to give metacognitive judgments increases, the calibration score should also increase; property # (2): if two students are equal in performance, the one with the more accurate metacognitive judgment should get the higher calibration score; property # (3): the calibration score

---

<sup>3</sup> These properties can also be used to determine the adequacy of measures of relative calibration (subsequent paragraph). However, these properties are introduced at this location to give the reader the opportunity to subsequently evaluate all introduced measures with regard to their adequacy as defined by Nelson (1984).

should be independent of performance on the criterion test; property # (4): if the ability to give metacognitive judgments is better for person A than for person B, the calibration score should be higher for person A than for person B, regardless of their performance on the criterion test and according to property # (5) also regardless of the overall difficulty of the criterion test; and property # (6): the calibration score should be independent of the type of criterion test (e.g., recognition or recall).

One of the most frequently used measures for capturing absolute calibration is the *bias* score (aka calibration bias or over- and underconfidence; Fitzgerald, White, & Gruppen, 2003; Maki & McGuire, 2002; Schraw, 1995; Schraw & Roedel, 1994; Stone, 2000; Winne & Jamieson-Noel, 2002): the mean signed arithmetic difference between predicted and observed performances; negative values represent underconfidence and positive values represent overconfidence. The second most frequently used measure is called *accuracy* (aka calibration accuracy, congruence, or deviation; Fitzgerald, White, & Gruppen, 2003; Maki & McGuire, 2002; Schraw, 1995; Winne & Jamieson-Noel, 2002): the mean unsigned difference between predicted and observed performances; it possesses only positive values which represent the magnitude of judgment error. Besides these rather simple measures, further coefficients of absolute accuracy have been proposed. For all these coefficients each item (each pair of metacognitive judgment and performance on the criterion test) is defined (a) as hit (predicted correct, performed correct), (b) as false alarm (predicted correct, performed incorrect), (c) as miss (predicted incorrect, performed correct) or (d) as correct rejection (predicted incorrect, performed correct). Thus, these coefficients compare the likelihood of correct and incorrect judgments. Schraw (1995), for example, argued for the *HC* (Hamann Coefficient:  $((a + d) - (b + c)) / ((a + d) + (b + c))$ ) with a range from -1 to +1 as a “reasonable measure of agreement accuracy” (p. 324).

Despite potential problems with coefficients of *absolute calibration* (e.g., margin dependence) especially Schraw (1995; Nietfeld, Enders, & Schraw, 2006; Schraw & Roedel, 1994) argued in favor of using such measures, at least in addition to measures of *relative calibration*. The most compelling argument for this demand is that relative calibration does not allow for conclusions with regard to absolute calibration because both kinds of calibration do not necessarily covary. To illustrate this argument consider the case where eight objects weigh 100, 110, 120, 130, 140, 150, 160, and 170 pounds. Judge A predicts weights of 130, 140, 120, 100, 170, 120, 130 and 140 pounds. Thus, she possesses an average *bias* of -4 (desirable) and judged and true weights correlate with  $r = .06$  (undesirable low). Judge B on the other hand predicts weight of 140, 150, 160, 170, 180, 190, 200 and 210 pounds. Thus she possesses an average *bias* of +40 (undesirable high) and judged and true weights correlate perfectly with  $r = 1.00$  (desirable; example taken from Schraw, 1995).

Unfortunately, the concept of absolute calibration can *hardly* be transferred to investigate the extended notion of calibration. This can be illustrated by an example: If students' calibration to the external demand *task complexity* should be investigated, measures of absolute



calibration would require a well-specified prescriptive model which details adequate task definitions, goals and plans and enacted strategies for all kinds of tasks. Such model would have to specify if the task “to memorize a list of unrelated nouns” requires a repetition strategy, an imagery strategy or any other kind of mixture between strategies. This kind of all-encompassing prescriptive model seems unrealistic. Even taxonomies that describe tasks in detail do not propose such a fixed set of strategies for task completion; rather they acknowledge that especially with more complex tasks a multitude of strategies can lead to successful task completion (e.g., Anderson et al., 2001).

However, to give an example of an empirical study using this research strategy (although the researchers never explicitly used the term absolute calibration): Winne and Jamieson-Noel (2002, 2003, Jamieson-Noel & Winne, 2003) had 69 students study a chapter on lightning with the software PrepMate. Students could choose one objective for learning: (1) describe important concepts, (2) explain cause and effect relations, (3) apply principles to explain phenomenon, or (4) generate and evaluate alternative solutions. After studying, students filled in a paper and pencil questionnaire on their use of study tactics. Participants’ self-reported study tactics were compared with a theoretical model about what kind of study tactics should be beneficial for which kind of learning objective. Results indicate that “students differentiated which tactics they reported using as a function of task” (Winne & Jamieson-Noel, 2003, p. 268), thus demonstrating their ability to *discriminate* (see previous section). However, the comparison with the theoretical model revealed that students in general reported overusing tactics with the simplest objective (describe important concepts) and in general reported underusing tactics with the most complex objective (generate and evaluate alternative solutions). Consequently, this study indicates low *absolute calibration*, probably because of an insufficient theoretical model.

### *Relative Calibration*

Within the traditional calibration paradigm, the discussion concerning adequate measures for relative calibration seems even more controversial. Most researchers agree, however, that measures of *association* or *correlation* relating metacognitive judgments and performance on the corresponding criterion test are adequate to capture relative calibration. A diverse set of such measures was suggested and promoted (for example, but not limited to):  $G$  (Goodman-Kruskal Gamma correlation for ordinal scales; Lin & Zabrocky, 1998; Maki & McGuire, 2002; Nelson, 1996; Nietfeld, Enders, & Schraw, 2006; Pressley & Ghatala, 1988; Schraw, 1995; Schraw & Roedel, 1994; Winne & Muis, 2004),  $r$  (Pearson’s product-moment correlation for interval scales; Fitzgerald, White, & Gruppen, 2003; Maki & McGuire, 2002; Schraw, 1995; Schraw & Roedel, 1994; Wiley, Griffin, & Thiede, 2005),  $d'$  ( $d$ -prime statistic from signal detection theory; Nelson, 1984; Winne & Muis, 2004), or  $V$  (Nelson’s formula for relative accuracy; Nelson, 1984, 1996).

Nelson (1984, 1996), for example, argued adamantly for the use of  $G$  (*Gamma correlation*; Goodman & Kruskal, 1954) as best measure to capture relative calibration, because it does not violate any of the above mentioned desirable properties and because of beneficial computational properties. For example,  $G$  is not margin dependent and does not require interval scaling.  $G$  provides a measure of association between two sets of scores analogous to a nonparametric correlation, ranging from -1 to +1. Originally, Schraw (1995) pointed out some shortcomings of  $G$  but those critiques were refuted by Nelson (1996).

This concept of relative calibration can be transferred to investigate the extended notion of calibration. For example, if students' calibration to the external demand *task complexity* should be investigated, measures of relative calibration would require a *systematic manipulation* of task complexity. Then, indicators of students' whole self-regulated learning process could be correlated with this property. For example, in the preparatory stage of learning relative calibration could be diagnosed if students considered more elaborate strategies like "evaluating critically" to be less important for simple tasks and of ascending importance for more complex tasks. To my knowledge, no empirical research explicitly and systematically addressing this question has been attempted so far, at least not for natural learning tasks.

However, Luwel, Verschaffel, Onghena, and De Corte (2003) applied this methodology to a very well-structured setting: Students were confronted with a 7 by 7 grid of squares in which randomly all possible fractions of numbers of squares were highlighted (49 trials). Students had to determine the correct number of highlighted squares as fast as possible. This whole process was repeated three times. In a choice condition (1) students could choose their strategy and the other two times they were forced to use one of two strategies: (2) the addition strategy (number of highlighted groups of squares are added) or (3) the subtraction strategy (from the total number of squares the number of non-highlighted squares is subtracted). For lower number of highlighted squares the addition strategy is most parsimonious, but if more than half squares are highlighted the subtraction strategy is best. Additionally, tasks with very few or almost all highlighted squares are simpler than tasks in which about half the squares are highlighted. Results indicate that in the (1) choice condition "participants indeed calibrated their strategy choice to the item characteristic" (Luwel, Verschaffel, Onghena, & De Corte, 2003, p. 525). The number of highlighted squares was negatively correlated with the percentage of participants that chose the addition strategy ( $r = -.92^{***}$ ) and the reverse was true for subtraction ( $r = .92^{***}$ ). Thus, participants demonstrated good *relative calibration* indicating that they in fact paid attention to task demands and adapted their strategy selection accordingly in this study. However, it is an open issue if this methodology can also be applied to more complex natural scenarios.

### *Calibration Graphs*

All previously introduced measures of calibration (*discrimination*, *absolute calibration*, and *relative calibration*) can be visualized by so-called calibration graphs. In the traditional calibration

paradigm these graphs illustrate the relationship between metacognitive judgments and performance on the criterion test (Figure 2.2-1). A calibration graph can be enriched with two supplementary aids: with a horizontal “line of nil discrimination” Weingardt, Leonesio, & Loftus, 1994) and with an ascending or descending “line of perfect calibration” (aka “utility line”, Stone, 2000) which represents a hypothetical clairvoyant with perfect *absolute calibration*. Weingardt, Leonesio and Loftus (1994) argue that this kind of representation is much better suited to represent calibration than any coefficients (1) because for each performance level students’ accuracy and over- or underconfidence can be immediately determined, (2) because they constitute a more detailed representation than coefficients of absolute or relative calibration, and (3) because the “line of perfect calibration” helps to evaluate students’ calibration with regard to an objective standard.

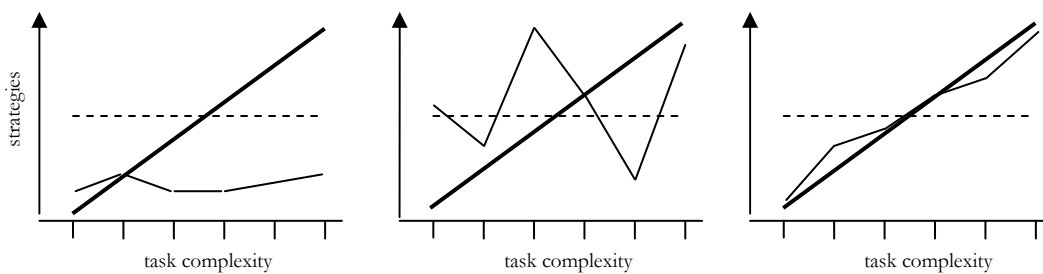


Figure 2.2-2: Hypothetical *calibration* graphs: Different levels of task complexity are plotted on the X-axis while students’ planned use of learning strategies is plotted on the Y-axis. The dashed horizontal line represents the “line of nil *discrimination*”. The dark ascending line represents the “line of perfect *calibration*”.

This concept of calibration graphs can be transferred to investigate the extended notion of calibration. For example, if students’ calibration to the external demand *task complexity* should be investigated in the preparatory stage of self-regulated learning, different levels of task complexity could be registered on the X-axis and students’ planned use of learning strategies could be registered on the Y-axis. If the measurements in a calibration graph are arranged in a horizontal line parallel to the “line of nil discrimination”, no *discrimination* can be diagnosed (Figure 2.2-2, left). On the other hand if the measurements deviate in the vertical dimension from a horizontal line, some *discrimination* can be diagnosed (Figure 2.2-2, middle and right). If the measurements in a calibration graph are close to the “line of perfect calibration”, high *absolute calibration* can be diagnosed (Figure 2.2-2, right). To diagnose relative calibration, measurements only have to be roughly parallel to the “line of perfect calibration” (without example in the Figure). To my knowledge, no empirical research explicitly and systematically addressing this question has been attempted so far.

### 2.2.4 Conclusion

In this chapter I highlighted several shortcomings of the traditional calibration paradigm<sup>4</sup> and argued for two extensions of the construct calibration in order to alleviate those problems and in order to enable the investigation of the core research questions of this thesis: (1) Not only learners' adaptation to *internal* conditions should be considered but also their adaptation to *external* conditions (extension of the criterion). (2) Not only learners' metacognitive *monitoring* should be considered, but also their active execution of metacognitive *control* strategies (extension of the subjective judgments). With these extensions, the core questions of this thesis can be re-conceptualized as questions of "calibration": How well-calibrated are students' learning processes with regard to external conditions and how are such calibration processes impacted by internal conditions. In other words: *the conditions and processes of metacognitive calibration* can be focused. Furthermore, I demonstrated that the methodology of the traditional calibration paradigm is mostly well-suited to investigate such questions of calibration in this new extended sense and thus should be used within this thesis. This is especially beneficial because these methods of the traditional calibration paradigm have not yet been systematically applied to questions of metacognitive adaptation in learning scenarios and thus this noticeable gap can also be interpreted as a desideratum for research. To conclude, the more general assumptions of the COPES-model about adaptation processes during learning can be supplemented with a more detailed conceptualization, terminology, and methodology derived from the traditional calibration paradigm.

As specified in the previous chapter, learners' calibration to two external conditions will be focused as core research question in this thesis: *task complexity* and *text complexity*. In order to apply the methods of the traditional calibration paradigm to this issue of adaptation, these dimensions need to be manipulated systematically. For example, tasks varying systematically in their objective complexity are needed (chapter 2.3). Then, learners' *discrimination* can be diagnosed by comparing their task definitions, goals, plans (in the *preparatory* stages), and their *enacted* strategies between tasks of different complexity (with repeated-measure ANOVAs). *Relative calibration* can be diagnosed by correlating task complexity with these constituents of the self-regulated learning process (e.g., with Kruskal-Goodman Gamma correlations). These measures could be further visualized by *calibration graphs*. By these means learners' adaptation to task complexity could be thoroughly scrutinized. Learners' adaptation to *text complexity* could be investigated analogously if texts of different objective complexity were created (chapter 2.4). In both cases *absolute calibration* can not be determined easily as no corresponding prescriptive models exist.

---

<sup>4</sup> Note, that the traditional calibration paradigm is considered adequate to capture learners' *internal monitoring processes* and therefore has a *raison d'être* in its own right. The critical review was just intended to highlight the limits of this methodology.

## 2.3 Task Complexity – An Important External Condition

*Task complexity* is one of the most influential external contextual demands that learners should adapt their self-regulated learning process to. This is not only assumed in the COPES-model (Winne & Hadwin, 1998; chapter 2.1), but also in other theoretical models: For example, Butler and Cartier's (2004) model of self-regulated learning posits that learners pay attention to three task features with the central underlying dimension of task complexity. Similar theoretical ideas are articulated in Broekkamp and van Hout-Wolters' (2006) model which considers task demands the most crucial contextual factor. Additionally, the importance of task complexity is obvious in everyday learning experiences: All learners are familiar with the phenomenon that some tasks are relatively easy to solve (e.g., "2 + 2"), while others are relatively hard to solve (e.g., an integral calculus equation).

However, it is less clear which factors exactly contribute to these differences in perceived task complexity. Thus, first some theoretical conceptualizations of task complexity will be discussed and potential systematic experimental manipulations will be considered in order to apply the methodology transferred from the traditional calibration paradigm (chapter 2.2). Empirical results reviewed subsequently consistently show that learners utilize different cognitive and metacognitive strategies to process tasks of different complexity and also reach different levels of understanding. This indicates that they might in fact monitor this external condition and adapt their self-regulated learning process accordingly.

### 2.3.1 Theoretical Conceptualizations and Potential Operationalisations

Most researchers conceptualize task complexity as an interaction between the *objective* task and the *subjective* learner. But it has to be emphasized that objective and subjective task complexity are not distinct entities, but more a matter of perspective: While subjective assumptions about complexity emphasize the perspective of the learner (Does the *learner* consider multiple procedures for task solutions?), objective assumptions state an objective perspective (Does the *task* afford multiple procedures for task solution?). To give an example: Imagine a mathematical word problem that requires a learner to infer an integral calculus equation and subsequently to solve it. Although this task has a fixed objective complexity the subjective complexity may vary tremendously, for example based on the learners' prior domain knowledge. While it might be solved almost automatically by a professional mathematician, a student just starting to learn about integral calculus might struggle to solve it at all. Thus, from a subjective perspective this task might be considered either very simple or very complex based on the learners' expertise.

### *Subjective Task Complexity*

With regard to subjective task complexity, two arbitrarily selected but representative theories will be introduced. First, Winne's (1996, 1997) idea about how learners derive their personal perceptions of task complexity will be reviewed. Winne posits that learners base their *ease-of-learning judgments* (EOLs) on three cues: (1) complexity, (2) speed and (3) reliability. Complexity (1) is assumed to depend on the number of cognitive operations needed to search memory, the number of IFs in a rule, and the number of rules. Thus, a task with a simple IF-THEN rule can be considered simple (IF two numbers are connected by a plus sign THEN add the numbers) while a task with complex IF-THEN-ELSE rules can be considered more complex. This definition of complexity is highly subjective because which kinds of rules need to be applied depends on the learners' prior domain knowledge. Speed (2) constitutes the time required for conducting a search or for executing a set of rules. This is also highly subjective because it also depends on the learners' expertise. Imagine a learner who looks at a familiar task and immediately knows how to solve that task. This learner estimates that task solution would be very fast and thus considers the task simple. On the other hand, imagine a learner who has no immediate idea about how to solve a task and has no experience in executing the appropriate operations. This learner would probably consider the same task complex. Reliability (3) can be defined as "the probability a tactic creates a product that meets preset standards" (Winne, 1997, p. 404). This constituent is also highly subjective for two reasons: First, because learners' standards can be assumed to be idiosyncratic, and second, because the learners' perception of their own abilities to reach those standards is also highly subjective ("self-efficacy"). Winne's idea of how learners derive EOLs can be summarized as follows: "cognitive tasks that involve one, fast, reliable cognitive operation are perceived to be easy" (Winne, 1996, p. 334), whereas all other tasks pose varying degrees of complexity.

A second theory that also focuses primarily on the subjective aspects of task complexity but also considers objective aspects is the *cognitive load theory* (CLT, Sweller, 2005; Sweller, van Merriënboer & Paas, 1998). CLT is primarily concerned with the instructional design of learning experiences considering the constraints of the human cognitive architecture. Nonetheless, some interesting insights about task complexity can be inferred. The central concern of CLT is working memory capacity: Only 7 +/- 2 chunks can be processed concurrently in short term memory (STM) (Miller, 1956; see also cognitive theory of multimedia learning, Mayer, 2001, 2005; Mayer & Moreno, 2003; integrated model of text picture comprehension, Schnotz, 2005; multi-component theory of working memory, Baddeley, 2003). In contrast to STM, long term memory (LTM) is assumed to be able to store unlimited amounts of information and is considered the locus of expertise which is assumed to be achievable via schema construction (Feltovich, Prietula, & Ericsson, 2006; Gruber, 2006). A schema can be anything that has been learned and is treated as a single entity (e.g., restaurant visit). Schemas reduce STM load as they can incorporate a huge amount of in-

formation but still represent only one chunk in STM. According to CLT, three different sources of cognitive load compete for STM resources during learning. When their total amount exceeds STM capacity, learning decreases. (1) The *intrinsic* cognitive load (element interactivity) depends on the number of elements (schemas) that must be simultaneously processed in STM during a task. It directly represents task complexity: In simple tasks only few elements have to be processed simultaneously while in more complex tasks element interactivity is higher. Element interactivity is highly subjective because what represents one element depends on the learners' prior domain knowledge. (2) *Extraneous* cognitive load is determined by instructional design (e.g., well-written text vs. demanding text; redundant material). Extraneous load also contributes to task complexity: Objectively task complexity can be varied by either writing a mathematical word problem in a very comprehensible way (simple) or by stating it like a fuzzy problem embedded in a real-life context (complex). (3) *Germane* cognitive load subsumes the conscious cognitive processes directly relevant for schema construction and represents the elaboration processes required for learning. To summarize, in CLT task complexity is conceptualized as addition of intrinsic cognitive load based on learners' expertise (subjective task complexity) and extraneous cognitive load based on the formulation of the task (objective task complexity).

Considering these theoretical ideas of *subjective task complexity* it is an open question how task complexity should be operationalised within the empirical studies of this thesis. All theoretical conceptualizations demonstrate that task complexity is highly dependent on learners' *prior domain knowledge*. According to these accounts it would be impossible to capture task complexity validly for all kinds of learners – independent of their prior domain knowledge. Thus, task complexity will not be operationalised subjectively.

#### *Objective Task Complexity*

Objective task complexity is often confused with task difficulty. Task *difficulty* is defined as the proportion of correct answers to a specific task within a sample (Bortz & Döring, 2002). Thus, task difficulty can vary within a fixed interval from 0 (most difficult task) to 1 (easiest task). For example, a task difficulty of .15 characterizes a difficult task as only 15 % of the learners were able to solve that task correctly. Task *complexity* on the other hand refers to cognitive complexity, for example, given by the level of cognitive processing. To further illustrate the difference, consider the subsequent examples: Factual questions (e.g., “What is the capital of Germany?”; correct answer: “Berlin”) require only one simple cognitive operation: searching for the answer in memory. Thus, this type of question possesses low task complexity. Still, this type of factual questions can be of a wide range of difficulties. For German citizens the question about the capital of Germany might be quite easy (probably around 100 % could answer it correctly). When the same type of question is asked with less familiar content (e.g., “What is the capital of Mongolia?”; correct answer: “Ulaanbaatar”) the task becomes far more difficult (probably only 20 % of Germans could

answer this question correctly) – but not more complex as the same cognitive operations have to be applied! The reverse case is also possible: Questions of the same difficulty might tremendously range in complexity. Thus, it can be inferred that *objective task complexity* is not the same as *task difficulty*, but rather that *objective task complexity* is closely related to *subjective task complexity* (see previous paragraph).

For defining *objective task complexity* two broad classes of theoretical conceptualizations have been suggested. The first class of conceptualizations can be described as cognitive process models derived from different situations such as decision making (Klayman, 1986) or information searching (Rouet & Tricot, 1996). These models are mostly grounded in cognitive psychology and consistently propose two common features that are assumed to differentiate tasks of different complexity: All such conceptualizations include the (1) *number* of elements (Campbell, 1988: multiple solutions, multiple goals; Klayman, 1986: number of available information; Rouet & Tricot, 1996: number of iterations of search cycles). Additionally, the multitude of goals, sequencing of these goals and the multitude of solution paths are also emphasized (Butler & Winne, 1995; Lodewyk & Winne, 2005; Perry, Philips, & Dowler, 2004). These conceptualizations are consistent with the theories described with regard to *subjective* task complexity (Sweller, van Marrienboer & Paas, 1998: element interactivity according to CLT; Winne, 1996, 1997: number of operations, number of IFs, number of rules). Tasks with more elements involved are consistently considered more complex. The *definitude or fuzziness* of a task (2) is not mentioned as consistently (Campbell, 1988: presence of uncertain or probabilistic links; Rouet & Tricot, 1996: fuzziness of goal). Tasks with more fuzziness (with fuzzy goals, erratic evaluations, moot standards, uncertainties, and potential conflict) are consistently considered more complex. This feature is related to the distinction between well-defined (simple) and ill-defined (complex) tasks.

The second class of conceptualizations can be described as task taxonomies which are mostly grounded in educational psychology and educational practice. These models consistently focus on a third feature that is assumed to primarily differentiate between tasks of different complexity: (3) the type of *knowledge* or type of *cognitive operations* required for task solution (Anderson et al., 2001; Biggs & Collis, 1982; Gagné, 1972; also chapter 2.3.2). Simple tasks are assumed to elicit superficial operations (e.g., memorizing) that focus on single concepts (e.g., names of animals). More complex tasks on the other hand are supposed to elicit a larger variety of more elaborate cognitive and metacognitive operations (e.g., integrating and critically evaluating content) which are assumed refer to larger chunks of knowledge (e.g., research about the evolution of animals).

Considering these theoretical ideas of *objective task complexity* it is an open question how task complexity should be operationalised within the empirical studies of this thesis. Operationalising task complexity by (1) the *number of involved elements* would be linked to problems: Conceptually, the number of operations is dependent on learners' prior domain knowledge and could not be defined independently. Thus, the same points of criticism as for subject-



tive conceptualizations of task complexity apply (see above). Operationalising task complexity by (2) the *definitude or fuzziness* alone would also not be satisfactory: This might be an important facet of task complexity, but does not tap the core construct. Contrary, the idea that (3) the type of necessary *cognitive operations* and the type of required *knowledge* determines task complexity can be considered more appealing and seems the best option to get valid results fairly independent of prior domain knowledge: Assuming that experts have primarily automated simple (memorize) and not complex cognitive operations (critically evaluate) experts – like novices – should perceive tasks which require more complex cognitive operations more complex. Additionally, this option is tentatively supported by empirical results: In study by van de Watering and van de Rijt (2006) 30 students were interviewed in detail about the criteria they applied when estimating task difficulty. Results indicate that high difficulty was especially assigned if the subject-matter was difficult per se (task required complex *cognitive operations* such as transfer; *knowledge* was accessed on a very detailed or a very abstract level) or if the task was phrased in a difficult way (task description contained multiple propositions). A further advantage of this option is practicability: Taxonomies exist that outline tasks of different complexity. For this thesis Bloom's revised taxonomy of educational objectives (Anderson et al., 2001) was selected as proxy for operationalising task complexity and will be reviewed in detail subsequently.

### 2.3.2 Bloom's (Revised) Taxonomy of Educational Objectives

The revised Taxonomy of Educational Objectives (subsequently referred to as "revised Taxonomy"; Anderson et al., 2001) will be considered as a central framework of this thesis. A defining characteristic of the term taxonomy is that the corresponding categories lie along a continuum. The revised Taxonomy possesses two dimensions: The organizing continuum underlying the most central cognitive process dimension is cognitive complexity (with the categories in order of ascending complexity: (1) *remember*, (2) *understand*, (3) *apply*, (4) *analyze*, (5) *evaluate* and (6) *create*). The revised Taxonomy was build to classify educational objectives which usually consist of a verb and a noun (Anderson et al., 2001). The verb generally describes the intended cognitive process (see above). The noun generally describes the type of knowledge required. Consider an example: "The student will remember (cognitive process) all states of the European Union (knowledge)." The organizing continuum underlying the second dimension, knowledge, reaches from concrete to abstract (with the categories in order of ascending abstraction: (1) *factual*, (2) *conceptual*, (3) *procedural* and (4) *metacognitive*). Both dimensions are combined into the Taxonomy Table which helps to classify all educational objectives (Table 2).

### *Bloom's Original Taxonomy of Educational Objectives*

To foster a deeper understanding of the cognitive process dimension, Bloom's original Taxonomy (subsequently referred to as "original Taxonomy"; Bloom et al., 1956), "one of the most influential educational monographs of the past half century" (Anderson & Sosniak, 1994, p. vii), will be reviewed first. It was developed by a large group of educationalists, psychologists and assessment specialists in the 1950ies (Anderson & Sosniak, 1994). This original Taxonomy is a uni-dimensional framework that enables classification of any objective in six categories of ascending complexity: (1) *knowledge*, (2) *comprehension*, (3) *application*, (4) *analysis*, (5) *synthesis*, and (6) *evaluation*.

More specifically, the category (1) *knowledge* consists of the sub-categories "*knowledge of specifics*", "*knowledge of terminology*", "*knowledge of specific facts*", "*knowledge of ways and means of dealing with specifics*", "*knowledge of conventions*", "*knowledge of trends and sequences*", "*knowledge of classifications and categories*", "*knowledge of criteria*", "*knowledge of methodology*", "*knowledge of the universals and abstractions in a field*", "*knowledge of principals and generalizations*", and "*knowledge of theories and structures*". These recall situations involve little more than bringing to mind the appropriate material (recall of previously learned material). All categories above *knowledge* are collectively labeled "abilities and skills" (Anderson et al., 2001). These more complex categories also possess sub-categories, but subsequently only the core definitions will be described: (2) *Comprehension* represents the lowest of these levels of understanding. "It refers to a type of understanding or apprehension such that the individual knows what is being communicated and can make use of the material or idea being communicated" (Anderson et al., 2001, p. 316). (3) *Application* constitutes "the use of abstractions in particular and concrete situations. The abstractions may be in the form of general ideas, rules of procedures, or generalized methods" (Anderson et al., 2001, p. 317). (4) *Analysis* implies "the breakdown [...] into its constituent elements or parts such that the relative hierarchy of ideas is made clear and / or the relations between the ideas expressed are made explicit" (Anderson et al., 2001, p. 317). (5) *Synthesis* involves "putting together of elements and parts so as to form a whole. This involves the process of working with pieces, parts, elements, etc., and arranging and combining them in such way as to constitute a pattern or structure not clearly there before" (Anderson et al., 2001, p. 318). (6) *Evaluation* encompasses "judgments about the value of material and methods for given purpose" (Anderson et al., 2001, p. 319).

### *Critical Aspects of Bloom's Original Taxonomy of Educational Objectives*

Bloom et al. (1956) claimed that the cognitive processes not only were ordered on a single dimension from simple to complex (*hierarchy assumption*), but also that "each class of behaviors was presumed to include all the behaviors of the less complex classes" (Kreitzer & Madaus, 1994, p. 66; *cumulative hierarchy assumption*). This means that all less complex behaviors have to be mastered before a more complex behavior can be mastered. On the other

hand, mastering a simple behavior is necessary but not sufficient to master more complex behaviors. These assumptions were criticized on a theoretical level and tested by several types of empirical studies.

On a *theoretical* level, these assumptions elicited fundamental logical and philosophical critique: For example, the claim that a single dimension might underlie all educational objectives has been considered too simplistic (Furst, 1994; Kreitzer & Madaus, 1994). Additionally, it can be assumed theoretically that categories of the original Taxonomy may overlap or inverse: “Indeed, for virtually any pair of taxonomic levels, a critic can be found who finds fault with the presumed order or even the separateness of those levels” (Kreitzer & Madaus, 1994, p. 67; also Furst, 1994).

A first type of *empirical studies* (Kreitzer & Madaus, 1994) involves judges categorizing assessment items to taxonomic levels. Those studies typically report the success of the categorization as interrater agreement. Results indicate that interrater agreement varies between 7 and 90 percent. Especially schoolteachers, the intended audience of the original Taxonomy failed to agree (Kreitzer & Madaus, 1994). For example, only for 33 out of 80 items teachers’ classifications were significantly better than chance. With better trained raters, higher interrater agreement could be achieved (up to 85 %). These empirical studies shed some light onto the communicability of the original Taxonomy.

A second type of *empirical studies* (Kreitzer & Madaus, 1994) tests “whether the test items aligned with successively more complex categories were successively more difficult” (Anderson et al., 2001, p. 288). Kropp and Stoker (1966, cited in Kreitzer & Madaus, 1994), for example, administered four domain-specific tests to over a thousand grade 9 through 12 students. Each test encompassed reading a text passage and answering 95 questions. Twenty items each tapped *knowledge*, *comprehension*, *application* and *analysis*, 5 tapped *synthesis* and 10 tapped *evaluation*. Consistent with the original Taxonomy’s expectations, students’ mean performance decreased for higher taxonomic levels. However, for one topic the performance on the *evaluation* subtest was better than on the *synthesis* subtest. For another topic, *application* proved easier than *comprehension*. Nonetheless, Kropp and Stoker themselves interpreted this result as support for the cumulative hierarchy. However, this type of research was critiqued as “fundamentally flawed” (Kreitzer & Madaus, 1994, p. 71) as task complexity and task difficulty are independent attributes that are not necessarily correlated (Guttman, 1953; Seddon, 1978; both cited in Kreitzer & Madaus, 1994).

A third type of *empirical studies* investigates the interrelationships among the six categories of the original Taxonomy. More specifically, students were administered items that represented specific categories of the original Taxonomy and their achievements for a given category were aggregated as a category score. These scores were analyzed using path analyses (Madaus, Woods, & Nutall, 1973), factor analyses, hierarchical syndrome analysis (Smith, 1968, cited in Kreitzer & Madaus, 1994), or structural linear equation modeling. For example, the above-mentioned data of Kropp and Stoker (1966, cited in Kreitzer &

Madaus, 1994) were re-examined to detect a simplex structure in the correlation matrices (Guttman, 1953; cited in Kreitzer & Madaus, 1994): For the simplest variable the correlation with the next complex variable should be highest, whereas the correlations with the more complex variables should decrease with increasing complexity of the second variable. Correlation matrices for each of the four subtests and each of the four grades were examined (a total of 16 correlation matrices). For one subtest all correlations conformed to the predicted pattern. For the other subtests the results were mixed. Thus, the researches concluded neither support nor rejection for the simplex pattern. If all studies in of this type of research are summarized, it can be concluded that “the three categories *comprehension*, *application*, and *analysis* were consistently ordered as predicted” (Kreitzer & Madaus, 1994, p. 72; also Anderson et al., 2001; Kreizer & Madaus, 1994). On the other hand, the proper placement of *knowledge* within this structure seems problematic (Kreitzer & Madaus, 1994). And for the most complex categories *synthesis* and *evaluation* the pattern was less clear (Smith, 1968; cited in Kreitzer & Madaus, 1994; Kreitzer & Madaus, 1994) but could be improved if the order of these categories was reversed (Anderson et al., 2001). Additionally, Madaus, Woods, and Nutall (1973) found that if general ability was included, all significant relationships between categories disappeared. This led the authors to conclude that assessment items at the more complex levels tap general ability more than specific content.

A fourth type of *empirical studies* investigates the assumption that solving more complex tasks should also lead to more complex cognitive processes and subsequently to more learning. The impact of task complexity on cognitive processes was demonstrated by Gierl (1997) who found that the “overall match between the response expected by the item writers and the response observed from the students [ $n = 30$ , responses determined by think aloud procedure] was 53.7 %” (p. 26). The impact of task complexity on learning outcome was investigated by Kunen, Cohen and Solman (1981) who presented 132 undergraduates with tasks of different categories of the original Taxonomy on the same content. Afterwards, students had to recall as many critical words as possible in a surprise recall test. Results indicate, that the lowest number of critical words was recalled after solving *knowledge* tasks, followed by *application* and *synthesis* tasks. Solving *evaluation* tasks did not lead to recalling the most critical words. Thus, the cumulative hierarchy assumption was validated for the categories *knowledge*, *application* and *synthesis*, but not for *evaluation*.

To summarize, tasks of the original Taxonomy seem to be hard to communicate to the intended audience (Kreitzer & Madaus, 1994), do not consistently elicit the intended cognitive processes (Gierl, 1997), and do also not consistently lead to intended learning outcomes (Kunen, Cohen, & Solman, 1981). The theoretically assumed cumulative hierarchy assumption was best validated for the mid-complex tasks *comprehension*, *application*, and *analyses*, while the *knowledge* category proved problematic as well as the most complex categories *synthesis* and *evaluation* (for a review see Anderson et al., 2001). Because of these empirical as well as theoretical critiques, Rohwer and Sloane (1994) conclude: “The structure claimed

for the hierarchy, then *resembles* a hierarchy, and the learning that makes possible the attainment of the objectives is *cumulative-like*” (p. 47, italics by the original authors). Seddon (1978; cited in Kreitzer & Madaus, 1994) added “it is perhaps fairest to say that the picture is uncertain. No one has been able to demonstrate that these properties do not exist. Conversely, no one has been able to demonstrate that they do.” (p. 321)

#### *The Revised Taxonomy of Educational Objectives*

Based on these and other empirical results and based on theoretical considerations, the original Taxonomy (Bloom et al., 1956) was recently revised (Anderson et al., 2001; also Krathwohl, 2002; Mayer, 2002). Changes in emphasis, terminology, and structure were implemented and will be detailed shortly.

The following represent the changes in *emphasis*: (1) The revised Taxonomy emphasizes the use of the Taxonomy in planning curriculum, instruction, and assessment opposed to the prime focus of the original Taxonomy on assessment (Airasian & Miranda, 2002). (2) The revised Taxonomy is aimed at a broader audience, especially at practicing teachers. (3) Only exemplary assessment items are provided in the revised Taxonomy. (4) Within the revised Taxonomy the six main categories are defined by extensive description of the sub-categories whereas in the original Taxonomy the six main categories were the main focus.

The following represent the changes in *terminology*: (1) Educational objectives indicate that a student should be able to do (verb) something (noun). Within the two-dimensional revised Taxonomy the knowledge dimension supplies the nouns and the cognitive process dimension supplies the verbs. Thus, the categories of the cognitive process dimension were labeled with verbs (e.g., *apply* instead of *application*). (2) The sub-categories were also expressed in verb forms. (3) The sub-categories of *knowledge* in the original Taxonomy were re-framed as unique dimension underlying each of the categories of the cognitive process dimension: *factual knowledge*, *conceptual knowledge*, *procedural knowledge*, and *metacognitive knowledge*. (4) *Comprehension* and *synthesis* in the original Taxonomy were re-titled (i.e., the former became *understand* and the later became *create*).

The following represent the changes in *structure*: (1) The noun and the verb components of objectives became separate dimensions: the knowledge dimension (*factual*, *conceptual*, *procedural*, and *metacognitive*) and the cognitive processes dimension (*remembers*, *understand*, *apply*, *analysis*, *evaluate*, and *create*). Therefore the verb aspect of the *knowledge* category in the original Taxonomy became *remember* in the revised Taxonomy as this verb most accurately captures what was implicitly implied in the original Taxonomy. (2) The two dimensions constitute the so-called Taxonomy Table, a four (knowledge categories) by six (cognitive processes categories) table that enables practitioners and researchers to sort all objectives and corresponding assessment items into the adequate cells (Table 2). (3) The cumulative hierarchy assumption of the cognitive process dimension was deleted: “An important characteristic of the revised Taxonomy, however, is that in order to conform to the language that

teachers use, the six categories are allowed to overlap on a scale of judged complexity” (Anderson et al., 2001, p. 309). Still, Anderson et al. (2001) argue that the revised Taxonomy can still be interpreted as a hierarchy, especially if the center of each category is considered: “The revised framework is a hierarchy in the sense that the six major categories of the cognitive process dimension are presumed to be ordered in terms of increasing complexity” (Anderson et al., 2001, p. 309). (4) Based on the better fit with the expectations of a cumulative hierarchy, the two most complex categories were reversed (*evaluation/evaluate* was considered less complex than *synthesis/create*).

For this project the *cognitive process dimension* of the revised Taxonomy is most important. Thus the subsequent paragraphs entail a detailed description (Anderson et al., 2001; Mayer, 2002; Krathwohl, 2002). The granularity of the differentiation can be given at the level of six main categories or at the level of nineteen sub-categories.

The cognitive process of *remembering* involves “retrieving relevant knowledge from long-term memory” (Mayer, 2002, p. 228; also Anderson et al., 2001). This is essential for all subsequent more complex cognitive processes: When meaningful learning is the goal (more complex cognitive processes) then *remembering* becomes a means to an end, rather than an end itself. *Remembering* encompasses two sub-categories: (1) Recognizing / identifying involves locating knowledge in long-term memory that is consistent with presented material. Verification, matching and forced choice tasks are adequate for assessing recognition. (2) Recalling / retrieving involves free recall. For assessment students are required to generate their own answers to open items.

Table 2: Taxonomy table of the revised Taxonomy (Table adapted from Anderson et al., 2001)

knowledge dimension	cognitive process dimension					
	remember	understand	apply	analyze	evaluate	create
factual						
conceptual						
procedural						
metacognitive						

The cognitive process of *understanding* constitutes the broadest category of the cognitive process dimension (Anderson et al., 2001; Mayer, 2002) with seven sub-categories: (1) interpret / clarify / paraphrase / represent / translate, (2) exemplify / illustrate / instantiate, (3) classify / categorize / subsume, (4) summarize / abstract / generalize, (5) infer / conclude / extrapolate / interpolate / predict, (6) compare / contrast / map / match, and (7) explain. Students are said to *understand* “when they are able to construct meaning from instructional messages” (Mayer, 2002, p. 228) and when can they integrate the incoming

knowledge with existing schemata and cognitive structures. For example, (2) exemplifying occurs when a student finds a specific example of a general principle and could, for example, be assessed by asking a student to pick a picture that exemplifies the impressionist painting style. Consequently, adequate assessment tasks for *understanding* (as well as for all subsequent more complex cognitive processes) require “that students cannot answer them correctly by relying on memory alone” (Anderson et al., 2001, p. 71). Assessment tasks for *understand* vary between sub-categories, but tasks can be either in the “constructed” response format (student is required to construct the response by herself) or in the “select” response format (student has to select the correct alternative from a number of presented alternatives, e.g., multiple-choice questions).

The cognitive process of *applying* involves “using procedures to perform exercises or solve problems” (Mayer, 2002, p. 229; also Anderson et al., 2001). This category is closely linked to *procedural knowledge* and encompasses two sub-categories: (1) Executing requires applying a procedure to a familiar task. To assess students’ ability to execute a newly acquired procedure, a math teacher could, for example, give them a worksheet with division exercises after having explained the appropriate procedure in class. (2) Implementing requires selecting and executing an adequate procedure for an unfamiliar task. For example, after introducing addition, subtraction, division and multiplication in math, a teacher could give her students a word problem which might be solved by one of these procedures. Thus, implementing requires more conceptual understanding than executing.

The cognitive process of *analyzing* involves “breaking material into its constituent parts and determining how the parts are related to each other and to an overall structure” (Mayer, 2002, p. 230; also Anderson et al., 2001). Thus, *analyzing* constitutes an extension of *understanding* or a prelude to *evaluating* or *creating*. This category has three sub-categories: (1) Differentiating / discriminating / selecting / distinguishing / focusing involves distinguishing relevant from irrelevant parts. It can be assessed by tasks that require selecting parts of presented material. (2) Organizing / finding coherence / integrating / outlining / parsing / structuring involves determining how elements function within a structure. Collecting and organizing pro and contra arguments constitutes an assessment task. (3) Attributing / deconstructing occurs by determining the point of view, biases, values, or intent in presented materials. To assess it a student could be asked to determine whether a psychological report was written by a behavioral or cognitive psychologist.

The cognitive process of *evaluating* involves “making judgments based on criteria and standards” (Mayer, 2002, p. 230; also Anderson et al., 2001). Most often, criteria refer to equality, effectiveness, efficiency or consistency. Sometimes, the criteria are externally provided, sometimes they may be determined by the student. *Evaluating* encompasses two sub-categories: (1) Checking / coordinating / detecting / monitoring / testing occurs when a student detects inconsistencies or fallacies within a process or product. Therefore, checking whether the conclusions stated within a research report are consistent with the data would

be an adequate assessment task. The second sub-category of (2) critiquing / judging occurs when a student detects inconsistencies between a product or operation and some external criteria. Therefore, *evaluating* two contradictory hypotheses with regard to their empirical support would be an adequate assessment task. The latter of these sub-categories is also at the core of what is also often referred to as critical thinking.

The cognitive process of *creating* involves “putting elements together to form a coherent or functional whole; that is, reorganize elements into a new pattern or structure” (Mayer, 2002, p. 231; also Anderson et al., 2001). This creative cognitive process involves three phases: (1) Within the divergent phase of problem representation (generating / hypothesizing) a variety of possible solutions are generated. Corresponding assessment requires the student to generate many possible solutions to a problem. (2) Within the convergent phase of solution planning (planning / designing) a solution method is devised and turned into a plan. Corresponding assessment requires describing a solution plan. (3) Within the third phase the solution plan is executed (producing / constructing). Corresponding assessment could require that students design a new product that meets certain specifications.

#### *Advantages of the Revised Taxonomy of Educational Objectives*

For the empirical studies of this thesis a classification system for *objective task complexity* is needed which has to fulfill the following criteria: (1) It must allow classification and ordering of all kinds of (learning) tasks with regard to their objective *cognitive* complexity, (2) the granularity of this classification should neither be too fine-grained nor too large-grained, and (3) as participants will be university students, the classification system should be able to differentiate cognitive processes at the higher end of the cognitive complexity continuum.

The cognitive process dimension of the revised Taxonomy fulfils all these conditions. Additionally, its hierarchy assumption was already empirically tested and can be considered a moderate fit with empirical data. Furthermore, the revised and original Taxonomies give ample examples on how to construct adequate assessment tasks on all levels of complexity. Another benefit concerns the fact that it is widely known in educational research and thus was used for fair number of empirical studies: For example, it was used as a framework to evaluate and give feedback on students’ writing (Athanassiou, McNett, & Harvey, 2003; Granello, 2001), to assess students’ metacognitive accuracy (Ayersman, 1995), or to construct adequate measures of students learning (Winne & Jamieson-Noel, 2002). The following paragraphs demonstrate that so far no better framework for the purpose of this thesis has been suggested. Thus, the choice of the revised Taxonomy is supported.

Alternative taxonomies and classification systems are numerous. Anderson et al. (2001) alone list eleven additional uni-dimensional frameworks (Ausubel & Robinson, 1969; Biggs & Collis, 1982; Bruce, 1981; Gagné, 1972; Gerlach & Sullivan, 1967; Hauenstein, 1998; Metfessel, Michael, & Kirsner, 1969; Quellmalz, 1987; Reigeluth & Moore, 1999; Romizowski, 1981; Stahl, 1979; all cited in Anderson et al., 2001) and eight additional mul-



tidimensional frameworks (DeCorte, 1973; Haladyna, 1997; Hannah & Michaelis, 1977; Marzano, 1992; Merrill, 1994; Ormell, 1974; Williams, 1977; all cited in Anderson et al., 2001). All of the listed multidimensional frameworks make knowledge and cognitive processes separate dimensions, but some of these frameworks also propose additional dimensions. All these frameworks were described and compared with the original (uni-dimensional frameworks) or revised (multi-dimensional frameworks) Taxonomy of Educational Objectives by Anderson et al. (2001).

In general, most of the listed uni-dimensional frameworks (see Anderson et al., 2001) are not able to differentiate more complex cognitive processes on a similar level of granularity as the original or the revised Taxonomy; instead they often concentrate on simpler cognitive processes. For example, the SOLO (Structure of Observed Learning Outcome) taxonomy of Biggs and Collis (1982) was conceptualized as a hierarchical model with five categories of ascending structural complexity: prestructural, unistructural, multistructural, relational and extended abstract. Anderson et al. (2001) argued that all categories map onto different sub-categories of the *knowledge* category of the original Taxonomy. Furthermore, extended abstract additionally also maps to the more complex categories of the original Taxonomy. To conclude, although this framework has a similar number of categories as the original Taxonomy, it differentiates better on the lower end of the cognitive complexity continuum while the original Taxonomy differentiates better on the higher end. Furthermore, empirical studies point to additional disadvantages of the SOLO taxonomy and no advantages over the original Taxonomy (Chan, Tsui, Chan, & Hong, 2002).

Most of the listed multi-dimensional frameworks (Anderson et al., 2001) also share problems with the uni-dimensional frameworks: most do not differentiate on a fine-grained level between more complex cognitive processes. Those that do (e.g., Ormell, 1974; cited in Anderson et al., 2001), correspond closely to the revised Taxonomy. Based on these strong similarities it is unlikely that they are superior. Some multi-dimensional classification systems that were not analyzed in by Anderson et al. (2001) will be shortly reviewed. Enokson (1973) developed a two-dimensional model that could be utilized by practicing teachers. The cognitive dimension encompasses the two categories high and low. Low is consistent with the original Taxonomy's conceptualization of the *knowledge* category, whereas all other of the original Taxonomy's categories were subsumed in high. The second dimension captures the nature of the task and the two categories are called convergent (closed tasks with one correct answer) and divergent (open tasks with many possibly correct answers). Therefore, this model – as well as others – is not able to differentiate more complex cognitive processes. Other multi-dimensional classification systems were devised for specific domains (e.g., Reid & Yang, 2002). De Jong and Ferguson-Hessler (1996), for example, proposed a model for physics. In a corresponding matrix six dimensions are considered: (1) type of knowledge (situational, conceptual, procedural, strategic), (2) level of knowledge (surface, deep), (3) structure of knowledge (isolated elements, structured knowledge),

(4) automation of knowledge (declarative, compiled), (5) modality of knowledge (verbal, pictorial), and (6) generality of knowledge (general, specific). These domain-specific classification systems also do not differentiate between more complex cognitive processes.

### 2.3.3 Empirical Results and Corresponding Theoretical Explanations

Empirical studies from as different research traditions as the traditional calibration paradigm, self-regulated learning, decision making and hypertext learning consistently show a high impact of *task complexity* on the cognitive and metacognitive components of the learning process as well as on the learning outcome.

#### *Results regarding Learning Strategies*

In a *self-report* study by Winne and Jamieson-Noel (2003) students reported more use of learning strategies for simple tasks than for complex tasks. This data is validated by studies concentrating on the *enactment* of strategies: For example, Veenman and Elshout (1999) had experts and novices of high and low intellectual ability solve problems of three levels of complexity. Only the expert group with high intellectual abilities consistently showed high metacognitive skills on all tasks. For all other groups, enacted metacognitive skills decreased with increasing task complexity. Similar results come from the *decision making paradigm*. For example, Klayman (1985) had 48 sixth grade students make four decisions of different complexity. To vary task complexity, the number of alternatives and number of evaluation dimensions was manipulated. The students had to explicitly ask for each piece of information and were instructed to stop if enough information for a good decision was gathered. Although the absolute number of information requests increased with ascending task complexity, it did not proportionally so. The percentage of information requested decreased from 79 % for the simplest tasks to 54 % for the most complex task. Similar results can be found in studies concerned with *hypertext learning*: For example, Rouet (2003) investigated the effects of task complexity (and prior domain knowledge) on hypertext searching. Twelve psychology students and 12 geography students worked on a hypertext on anorexia and a hypertext on Peru. For each text, 2 simple specific questions and 2 complex general questions had to be answered. Results indicate fast and precise searches for simple tasks and longer and less precise searches for more complex tasks.

#### *Results regarding the Learning Outcome*

Empirical studies not only consider the learning process, but also the learning outcome. For example, Gall (2006) had 23 students from the US Air Force Academy learn with a hypertext about the development of airpower during WWI. Based on a classification system by Gall and Hannafin (1994), learners were randomly assigned one of three conditions

where they were asked to “browse” (explore the hypertext, simple task), “search” (list of factual questions, moderate task) or “connect” (list of conceptual questions, complex task). Logfiles indicate that the “searching” condition elicited a high number of shortly accessed hypertext nodes while the “connecting” condition elicited fewer but longer node accesses. In a posttest all learners had to complete 10 factual and 2 conceptual tasks. For both kinds of questions the “connecting” group significantly outperformed the “searching” group which in turn outperformed the “browsing” group.

#### *Results regarding Calibration*

Within the traditional calibration paradigm better calibration is mostly associated with simpler tasks (Burson, Klayman, & Larrik, 2006; Pressley & Ghatala, 1988; Schraw & Roedel, 1994; Schraw, Potenza, & Nebelsick-Gullet, 1993). Schraw and Roedel (1994), for example, administered a reading comprehension test with easy (70 – 90 % correct in a previous study), moderate (50 – 70 % correct) and difficult (30 – 50 % correct) tasks. After solving each task, students ( $n = 48$ ) gave confidence ratings for their performance. Results show underconfidence for easy and moderate tasks and overconfidence for difficult tasks. The absolute bias was greater for difficult tasks than for either the moderate or easy tasks. These results are congruent with a study by Pressley and Ghatala (1988) who found that calibration was consistently correlated with task difficulty: Students ( $n = 51$ ) were more aware of performance for easier tasks (better calibrated).

#### *Summary of Empirical Results*

For *simple* learning tasks, students demonstrate an effective self-regulated learning process by showing superior calibration, better learning strategies, and more complete information search (Bienmiller, Shaney, Inglis, & Meichenbaum, 1998; Klayman, 1985; Payne, 1976; Pressley & Ghatala, 1988; Rouet, 2003; Rouet, Vidal-Abarca, Erbou, & Millogo, 2001; Schraw & Roedel, 1994; Veenman & Elshout, 1999; Winne & Jamieson-Noel, 2002) as well as good recall of facts (learning outcome: Brunstein & Krems, 2005). However, note that such simple tasks seem to be detrimental for achieving deeper understanding of the subject matter (Gall, 2006). Additionally, simple tasks seem to elicit a long-term superficial view of learning: For example, in Eisenberger’s studies (Eisenberger, 1992; Eisenberger, Masterson, & McDermitt, 1982) students were presented with multiple tasks that varied in content as well as in complexity (adding 2-digit or 7-digit numbers). Students who received tasks from all content areas and who received the tasks that required most effort (complex tasks) did best on a transfer test. Thus, the reverse effect is also feasible: Students who are constantly confronted with very simple tasks will probably not be able to engage in adequate deep processing if confronted with a more complex task. Thus, it is not surprising that educators demand that schools and universities should use a variety of tasks, especially more complex ones (Anderson et al., 2001; Lodewyk & Winne, 2005; Perry, Philips, & Dowler, 2004).

For more *complex* learning tasks, students demonstrate a less adequate self-regulated learning process by showing less calibration, fewer learning strategies, and fragmentary information search (Bienmiller, Shaney, Inglis, & Meichenbaum, 1998; Klayman, 1985; Payne, 1976; Pressley & Ghatala, 1988; Rouet, 2003; Schraw & Roedel, 1994; Veenman & Elshout, 1999; Winne & Jamieson-Noel, 2002), but superior conceptual understanding of the subject matter (learning outcome: Brunstein & Krems, 2005; Gall, 2006).

#### *Potential Theoretical Explanations*

At least two explanations for this pattern of results are feasible: First, Klayman (1985) proposed the idea of an *effort / accuracy trade-off*. According to this hypothesis, decision makers adapt their information search process to task complexity, but they are not willing to “pay the full cost”. Instead, they compromise with regard to accuracy to spare some additional effort required for an optimal decision. This explanation indirectly claims that all students have the capacity to reach an optimal solution – even for complex tasks – but lack the will.

Second, Lin and Zabrocky (1998) proposed the idea of *limited cognitive capacity* consistent with cognitive load theory (Sweller, van Merriënboer, & Paas, 1998). This hypothesis entails that “easy tasks require less cognitive effort and less intensive processing” (Lin & Zabrocky, 1998; p. 377). Thus, for simple tasks the free cognitive resources can be used for additional self-regulation. With more complex tasks on the other hand, the cognitive capacity is exhausted by the cognitive processes alone and no capacity is left for additional self-regulation. This explanation claims that students – especially poor students – may lack the capacity to reach an optimal solution for complex tasks even though they may be willing.

The latter explanation receives support from empirical studies demonstrating that cognitive and metacognitive processes in fact compete for working memory capacity: Bienmiller, Shany, Inglis, and Meichenbaum (1998) showed that all students are capable of demonstrating self-regulation “if we make tasks easy enough and give them a reason to talk” (p. 214), but not under more complex task conditions. Kanfer and Ackerman (1989) showed that Air Force personnel learning the complex skill of air traffic controller could either optimize their performance (cognitive process) or their self-regulation (metacognition), but not both.

#### **2.3.4 Conclusion**

To investigate the conditions and processes of metacognitive *calibration to task complexity*, this external condition needs to be systematically manipulated to constitute an adequate independent variable. The selective review of corresponding theoretical conceptualizations shows that task complexity is highly dependent on prior domain knowledge. Therefore, it was hard to find an adequately objective operationalisation. I decided to focus on the underlying *cognitive processes* and chose Bloom’s revised Taxonomy (Anderson et al., 2001) as

best-suited framework. Consequently, in the empirical studies of this thesis task complexity will be systematically manipulated by constructing tasks of the six categories of Bloom's revised Taxonomy (subsequently called *Bloom-Categories*; in order of ascending cognitive complexity): (1) *remember*, (2) *understand*, (3) *apply*, (4) *analyze*, (5) *evaluate*, and (6) *create*. Note that these categories are assumed to form a hierarchy. Therefore, it will be possible to determine *relative calibration* by rank correlations (Goodman-Kruskall's Gamma,  $G$ ) between these *Bloom-Categories* and learners' planned or enacted self-regulated learning processes.

Based on the reviewed empirical findings and theoretical explanations the following effects of task complexity as independent variable can be expected in the empirical studies of this thesis: First, learners will adapt their cognitive and metacognitive learning strategies to task complexity, for example by using strategies of deeper elaboration for more complex texts. Second, learners will demonstrate better learning outcome for simple tasks. And third, these adaptation processes will potentially interact with internal conditions, for example with learners' expertise or intellectual ability (Veenman & Elshout, 1999).

## 2.4 Text Complexity – An Important External Condition

Text complexity is another influential external contextual demand that learners should adapt their self-regulated learning process to. This is not only assumed in the COPES-model (Winne & Hadwin, 1998; chapter 2.1), but also in other theoretical models: For example, Reynolds (1992) elaborates that readers who apply the selective attention strategy (SAS) consider three cues: characteristics of the text, characteristics of the reader, and characteristics of the task. One important text characteristic is text complexity, for example defined as cohesion or information density. Additionally, the importance of text complexity is obvious in everyday learning experiences: All readers experience the phenomenon that some texts are relatively easy to comprehend (e.g., children's fairytales), while others are relatively hard to comprehend (e.g., Kant's Critique of Pure Reason).

However, it is less clear which factors exactly contribute to these differences in perceived text complexity. Thus, first some theoretical conceptualizations of text complexity will be discussed and potential systematic experimental manipulations will be considered in order to apply the methodology transferred from the traditional calibration paradigm (chapter 2.2). Empirical results reviewed subsequently consistently show that readers utilize different reading and learning strategies to process texts of different complexity and also reach different levels of understanding. This indicates that they might in fact monitor this external condition and adapt their self-regulated learning process accordingly.

### 2.4.1 Theoretical Conceptualizations and Potential Operationalisations

In order to answer the question why some texts are perceived as simple and others as more complex, it is helpful to shortly review theoretical conceptualizations of text comprehension. Snow (2002; also McNamara, Louwerse, & Graesser, 2007), for example, assumes that the text comprehension process includes four interactive components: characteristics of the reader (e.g., prior knowledge about the content), the text (e.g., the exact words and sentences), the comprehension activities (e.g., the reading strategies), and the sociocultural context. Based on these constituents the reader may construct different levels of text representations (Kintsch, 1998): A *surface* representation entails recall of exact wording and syntax, a *textbase* representation entails recall of coherent propositions that preserve the meaning of a text (but not the exact wording and syntax), and a *situation model* representation entails a microworld which represents the content of the text (i.e., constructed by inferential interaction between the explicit text, background knowledge of the reader, and the comprehension goals of the reader). In most learning scenarios readers are required to attempt to build a *semantic* text representation (*textbase* or *situation model*) which also depends on reader characteristics like prior domain knowledge and reading strategies. Consequently, these models of text comprehension highlight that in most learning scenarios the complexity of a specific text can hardly be diagnosed by looking at the text alone. Rather *text complexity* is a multidimensional construct with a focus on the interaction between a text and a reader with *objective* as well as *subjective* properties (Ballstaedt, 1997).

#### *Objective Text Complexity*

The availability of about 200 *readability formulas* (Duran, Bellissens, Taylor, & McNamara, 2007) attest to the continuing attempts to operationalise *objective* text complexity or text comprehensibility. The most widely known examples of such formulas are the Flesch-Kincaid Grade Level (Flesch, 1948) that indicates the years of education supposedly required to comprehend a given document and the Flesch Reading Ease Score (Flesch, 1948) that indicates the difficulty of comprehending a document on a scale from 0 (difficult) to 100 (simple). These formulas use text properties such as word length and sentence length as indicators of difficulty and have clear limitations. For example, a simple text according to these indices may be constructed from short and choppy sentences which may make it difficult to comprehend. This problem may arise due to deficient text cohesion because explicit linguistic elements that link sentences might be missing (McNamara, Louwerse, & Graesser, 2007). Furthermore, an extremely difficult content on a semantic level (*situation model*) might be described with short words in short sentences (be objectively easy according to the readability formulas but be subjectively difficult to comprehend). This problem arises due to the fact that simple readability scores *only* concentrate on *surface* features of the text and neglect the content. Even the most sophisticated readability formulas that try to

model the interaction between text and reader by taking into account lexica of known words (Duran, Bellissens, Taylor, & McNamara, 2007, McNamara, Louwerse, & Graesser, 2007: indices created by Coh-Metrix; Landauer, Foltz, & Laham, 1998: LSA, latent semantic analysis) are not yet capable of determining text complexity in a satisfactory way (apart from the problem that these methods were developed for the English language and cannot be transferred easily to other languages).

Besides these readability formulas, another attempt to capture objective text difficulty focuses more strongly on semantic content: the concept of *information density* (Ballstaedt, 1997; Kintsch, 1998). Within this method the number of propositions per specific unit (e.g., 100 words) is determined. Texts with a higher information density are considered more difficult. This method has the advantage of concentrating at least on a *textbase* representation of the text (Kintsch, 1998) and not only on surface features. However, besides the high cost (effort in determining the propositions), this method almost neglects the surface representation of the text. To highlight this problem, consider that a specific content (*situation model*) might be easy or difficult to comprehend based on the way it is described in writing (with long words and nested sentences vs. with short words and simple sentences). Thus, the surface features of a text should not be ignored.

Neither readability formulas nor the concept of information density seem well-suited as stand-alone options to operationalise *text complexity objectively*. The former ignores semantic representations while the latter ignores surface representations. However, these theoretical ideas could be used to facilitate construction of differently complex texts: Complexity should mainly be manipulated at deeper levels of text representation (*textbase* or *situation model*; Kintsch, 1998) and not only on the *surface* level. Therefore, texts of different complexity will be constructed mainly by describing differently complex situation models.

### *Subjective Text Complexity*

Opposed to such attempts to operationalise text complexity as objectively as possible, text complexity or text understanding can also be captured *subjectively* (Jucks, 2001). Groeben (1972), for example, proposed – based on a theoretical analysis – four dimensions that are assumed to differentiate between simple and more complex texts: (1) “stylistic simplicity” refers to the benefit of short sentences with few subordinate clauses that use active verbs, (2) “semantic redundancy” refers to the fact that texts without literal repetitions and voluminousness are more comprehensible, (3) “cognitive structuredness” refers to the positive impact of advanced organizers, emphasis of important concepts, or summaries, and (4) “conceptual conflict” refers to the fact that the depth of processing can be increased by adding incongruent concepts or questions to the reader. An alternative four-dimensional framework based on empirical studies was suggested by Langer and colleagues (Langer, Schulz von Thun, & Tausch, 1990): (1) “simplicity” in word choice and short sentences, (2) “structuredness” of the text, and (3) “conciseness” contribute to text comprehensibility,

whereas (4) “stimulating additions” enhance the readers’ motivation. To apply these subjective methods readers have to *evaluate* a text with these predefined dimensions. These methods clearly take into account the interaction between the reader and the text, the neglect of which can be considered problematic for most objective methods described above. Furthermore, these holistic judgments most likely consider *surface* features of the text such as words and syntax (“stylistic simplicity”, Groeben, 1972) as well as semantic *textbase* features such as the cohesion of the propositions of a text (“structuredness”, Langer, Schulz von Thun, & Tausch, 1990). However, consider two disadvantages: First, these methods “cost” more than most objective methods because they require human judges; and second, these methods also do not explicitly refer to the *situation model* of a text (e.g., the consistency with a learner’s prior domain knowledge, Kintsch, 1998).

These theoretical considerations as well as empirical data tentatively indicate that *text complexity* could be operationalised *subjectively*: Reader judgments would be best suited to validate the differential complexity of texts. Therefore, pilot studies will ask readers to give holistic judgments about the complexity and comprehensibility of newly constructed texts (see section on objective text complexity). This option is also tentatively supported by empirical data: Daghighi, Fattori and Ciardullo (2006), for example, compared the Italian version of the Flesch Reading Ease Score with self-report measures about how difficult a text was to read and how many words the readers did not know. Furthermore, comprehension was objectively assessed by a knowledge test. While readability scores indicated high text complexity, more subjective measures indicated low text complexity: Even laypersons reported high readability, few unknown words, and were able to answer the knowledge questions successfully. Similar results were obtained by Quereshi (1989).

## 2.4.2 Empirical Results and Corresponding Theoretical Explanations

Empirical studies from as different research traditions as the calibration paradigm, self-regulated learning, and text comprehension consistently show a high impact of *text complexity* on the cognitive and metacognitive components of the learning process as well as on the learning outcome.

### *Results regarding Learning Strategies and Learning Outcome*

To investigate learning strategies and learning outcome, Veenman and Beishuizen (2004) captured readers’ online reading activities. They presented 46 university students with one moderately complex and one complex text which had to be studied either under time pressure or not. Afterwards, students’ learning outcome was captured by a multiple-choice test. Think aloud protocols revealed that the more complex text elicited learning strategies of deeper elaboration. While no difference in performance could be detected for the moderate



text, time pressure was detrimental to the learning outcome for the complex text. Empirical studies by Kintsch and colleagues focus primarily on readers' level of understanding (learning outcome). For example, in a study by McNamara, Kintsch, Songer, and Kintsch (1996, 2<sup>nd</sup> experiment) students ( $n = 56$ ) were presented with one of four texts, ranging from a maximal coherent text (coherent at the local and the global level) to a minimally coherent text (neither coherent at the local nor at the global level). Results indicate that text processing and learning not only depend on text complexity but also on students' prior domain knowledge. Readers with low prior domain knowledge benefited from coherent texts. High prior knowledge readers on the other hand benefited from a minimally coherent text, at least with regard to measures of deep comprehension.

To summarize, readers tend to use strategies of deeper elaboration for more complex texts (Kim, 2003; Veenman & Beishuizen, 2004), but text complexity may be detrimental for learning outcome, at least for less knowledgeable readers (also Maki, Shields, Wheeler, & Zacchilli, 2005; McNamara, Kintsch, Songer, & Kintsch, 1996; Salmerón, Kintsch, & Canas, 2006; Weaver & Bryant, 1995).

#### *Results regarding Calibration*

Within the traditional calibration paradigm calibration is not consistently associated with text complexity (Lin & Zabrocky, 1998; Magliano, Little, & Graesser, 1993; Maki, Foley, Kajer, Thompson, & Willert, 1990; Maki, Shields, Wheeler, & Zacchilli, 2005; Rawson, Dunlosky, & Thiede, 2000; Schommer & Surber, 1986; Weaver & Bryant, 1995). Maki et al. (1990) demonstrated that calibration for texts with deleted letters was better than calibration for intact texts. Weaver and Bryant (1995) on the other hand found that calibration was best for a moderate level of text difficulty ( $G$  around .70) compared to easy ( $G$  around .30) and difficult texts ( $G$  around .30). Maki, Shields, Wheeler, and Zacchilli (2005) investigated the same question on a more detailed level: Measures of absolute calibration demonstrate that for difficult texts low ability students showed overconfidence in predictions whereas medium and high ability students showed accurate predictions, but underconfidence in postdictions. With regard to relative calibration, no significant difference between texts of different complexity could be detected (although the calibration indices for simple texts,  $G = .20$  for predictions and  $G = .36$  for postdictions, were generally lower than those for difficult texts,  $G = .36$  for predictions and  $G = .57$  for postdictions). To summarize, most results indicate that calibration might be better for texts of at least moderate difficulty than for simple texts (Maki et al., 1990: deleted letter text; Maki et al., 2005: relative calibration; Weaver & Bryant, 1995: standard texts). However, some empirical results also point in the opposite direction (Weaver & Bryant, 1995: difficult texts).

### *Potential Theoretical Explanations*

In order to explain these empirical effects, Lin and Zabrocky (1998) as well as Weaver and Bryant (1995) proposed the “optimum effort hypothesis”: Very *simple* texts are assumed to elicit automated surface processing and thus conscious metacognitive processes necessary for calibration – or for strategies of elaboration – might not be enacted (Maki et al., 1990: intact texts; McNamara, Kintsch, Songer, & Kintsch, 1996: coherent texts for knowledgeable readers; Schommer & Surber, 1986: easy texts; Weaver & Bryant, 1995: easy texts; Veenman & Beishuizen, 2004: moderate text). Texts of *moderate* complexity on the other hand are assumed to require students’ conscious – and thus deeper – processing but also leave enough cognitive resources for metacognitive processes, thus resulting in “optimum effort” (Maki et al., 1990: deleted letter texts; McNamara, Kintsch, Songer, & Kintsch, 1996: coherent texts for the less knowledgeable readers, incoherent texts for the knowledgeable readers; Weaver & Bryant, 1995: moderately difficult texts; Veenman & Beishuizen, 2004: difficult text). Very *complex* texts are assumed to afford very deep processing and thus leave few resources for metacognitive processes; even if the learning process might be metacognitively regulated, optimal performance in general is not obtained (McNamara, Kintsch, Songer, & Kintsch, 1996: incoherent texts for less knowledgeable readers; Weaver & Bryant, 1995: difficult texts). This explanation is consistent with the assumption of a limited capacity of working memory as proposed by the cognitive load theory (Sweller, van Merriënboer, & Paas, 1998).

Furthermore, empirical results demonstrate that *internal* as well as *external conditions* as defined by the COPES-model (Winne & Hadwin, 1998) are more relevant for comprehension of more complex texts: For example, the impact of time constraints (Veenman & Beishuizen, 2004), learners’ prior knowledge (McNamara, Kintsch, Songer, & Kintsch, 1996) or their reading goals (Schommer & Surber, 1986). Wiley, Griffin and Thiede (2005) propose an explanation based on Kintsch’s model of text comprehension (1998; Van Dijk & Kintsch, 1983). They argue that for simple texts, the *surface* representation, the *textbase* and the *situation model* would be quite similar. Thus, deep comprehension cannot be distinguished from pure recall of text elements. In more complex texts on the other hand, these elements may differ. Thus, “only complex texts provide researchers the opportunity to test for situation model level inferences that cannot be simply recalled from the text-base” (Wiley, Griffin, & Thiede, 2005, p. 412).

### **2.4.3 Conclusion**

To investigate the conditions and processes of metacognitive *calibration to text complexity*, this external condition needs to be systematically manipulated to constitute an adequate independent variable. The selective review of corresponding theoretical conceptualizations

shows that so far no easy-to-use operationalisation of text complexity exists. *Objective* as well as *subjective* methods both have their specific merits and shortcomings. Thus, based on multiple arguments (see above) I decided to use a combination of methods: Texts of different complexity will be created by describing differently complex *situation models* (Kintsch, 1998) with different numbers of *propositions* per information unit (*objective* conceptualization). In anticipation of the experimental design it should be noted that clusters of texts of similar complexity will be the focus of analysis. More specifically, a hypertext with three hierarchical levels will be constructed (in order of ascending complexity): (1) “*level 1*” (simple texts that serve as introductions), (2) “*level 2*” (moderately complex texts that elaborate content), and (3) “*level 2*” (complex texts that provide very detailed information about certain aspects of the content). In a second step, these newly created texts will be piloted by having readers evaluate them on several dimensions to validate their differential complexity (*subjective* conceptualization). As these hypertext levels form a hierarchy it will be possible to determine *relative calibration* by rank correlations (Goodman-Kruskall’s Gamma,  $G$ ) between hypertext levels and learners’ self-regulated learning processes.

Based on the reviewed empirical findings and theoretical explanations the following effects can be expected in the empirical studies of this thesis: First, learners will adapt their cognitive and metacognitive learning strategies to text complexity, for example by using strategies of deeper comprehension for more complex texts. Second, these adaptation processes will interact with internal conditions: Learner characteristics will play an increasingly important role for more complex texts.

## 2.5 Prior Domain Knowledge – An Important Internal Condition

Prior domain knowledge is one of the most influential learner characteristics and should impact the whole self-regulated learning process. Within the COPEs-model (Winne & Hadwin, 1998) it is assumed that learners “draw on knowledge [...] to construct an interpretation of a task’s properties and requirements” (Butler & Winne, 1995, p. 248). Furthermore, Winne and colleagues argue that “for experts, control is embedded within domain knowledge. Thus, needs to metacognitively monitor and control task engagement are minimal” (Winne, 1996, p. 333). This significant impact of prior domain knowledge is endorsed by other theoretical models: For example, Borkowski, Chan, and Muthukrishna (2000) claim that “knowledge is often sufficient to solve problems, even without the aid of strategies. In these situations metacognitive processes [...] are unnecessary.” (p. 9). Additionally, the importance of prior domain knowledge is obvious in everyday learning experiences: An expert is able to solve most tasks within her domain easily, for example due to her extensive knowledge of technical terms and due to her automated skills attained by extended practice. A novice on the other hand has to laboriously solve these tasks in a

step-by-step fashion. As an example consider having to design your own house without knowledge in architecture.

The expert paradigm entails corresponding theoretical models as well as empirical research investigating this phenomenon. Thus, first theoretical conceptualizations of expertise will be discussed and potential operationalisations will be considered. Empirical results reviewed subsequently demonstrate significant impact of prior domain knowledge on most aspects of learners' metacognitively governed self-regulated learning process.

### 2.5.1 Theoretical Framework – the Expert Paradigm

The expert paradigm is the most pervasive theoretical framework of prior knowledge differences. This paradigm tries to answer the fundamental question of how expertise develops and how it can be trained. Expert-novice comparisons are the preferred methodology; sometimes levels of intermediate expertise are included. Throughout history multiple theories about expert abilities were proposed within this paradigm – and often rejected.

#### *General-Process View*

Starting in the 16<sup>th</sup> century, earliest explanations for superior expert talents were based on the concept of divine god-given gifts (Ericsson, 2005). This idea was replaced by hereditary explanations in the 19<sup>th</sup> century: The capacity for expertise was supposed to be passed on from parents to children (Ericsson, 2005). The first psychological accounts of expertise were consistent with this idea. Experts were hypothesized to possess superior information processing capacities or better domain-general problem solving skills (Feltovich, Prietula, & Ericsson, 2006; Gruber, 1994; Gruber & Ziegler, 1996). This *general-process* view of expertise (Sternberg, 1997) inspired, for example, research in artificial intelligence (AI) trying to implement general computational algorithms (weak methods; Feltovich, Prietula, & Ericsson, 2006). If this *general-process* view was correct, experts should excel in multiple domains and should be able to transfer their superiority at least related domains (Gruber, 1994). This view was disproved by Chase and Simon (1973) who demonstrated that the superior memory of chess experts depended on their superior *domain-specific knowledge* and not on their general superior cognitive capacity. Experts only demonstrated superior memory for meaningful chess positions, but not for scrambled chess boards (for an overview of similar results see Chi, Glaser, & Rees, 1982; Gruber, 1994).

#### *Quantity-of-Knowledge View*

Consequently, the historically subsequent explanation for experts' superior abilities was based on the idea that experts possess quantitatively more domain-specific knowledge. This *quantity-of-knowledge* view (Sternberg, 1997) also inspired, for example, research in artificial

intelligence (AI) trying to implement so-called strong methods (Feltovich, Prietula, & Ericsson, 2006; Gruber, 1994). In order to research this domain-dependent expertise, domain-specific tasks were used, for example from geometry, physics, thermodynamic, programming or electronics. Empirical results support this *quantity-of-knowledge* view (for an overview see Chi, 2006): Experts not only demonstrate superior memory for meaningful stimuli (see above), but also possess quantitatively more domain-specific knowledge (Bromme, 1992), make fewer mistakes in problem solving (Chi, Glaser, & Rees, 1982; Feltovich, Prietula, & Ericsson, 2006; Gruber & Strube, 1989; Johnson et al., 1981; Larkin, McDermott, Simon, & Simon, 1980) and solve problems faster (Chi, Glaser, & Rees, 1982; Feltovich, Prietula, & Ericsson, 2006; Gruber & Strube, 1989; Larkin, McDermott, Simon, & Simon, 1980; Rothe & Schindler, 1996).

A consistent theoretical explanation was proposed by the *chunking theory* of Chase and Simon (1973; also Ericsson, 2005; Gobet, 2006). This theory assumes that experts and novices are subjected to the same memory constraints and limitations (human information processing theory, Newell & Simon, 1972), for example human short term memory is assumed to be limited in capacity to 7 +/- 2 chunks of information (Miller, 1956). But furthermore the chunking theory posits that experts are better adapted to task constraints and are better able to circumvent their cognitive limitations because they save larger meaningful chunks of information in long term memory. In chess, for example, novices have to memorize the position of each chessman separately resulting in approximately 7 memorized positions. For experts on the other hand 7 chessmen might represent one chunk if they build a meaningful formation. According to this theory, chunking is the main mechanism of expertise development.

If this *quantity-of-knowledge* view was correct, everyone should be able to gain expert status by accumulating more domain specific knowledge, independent of intellectual abilities (Mack, 1996). Furthermore, performance should increase inevitably due to experience (Ericsson, 2005). This view of expertise was refuted, for example, by Chi, Feltovich and Glaser (1981, see subsequent section) and by Larkin, McDermott, Simon and Simon (1980). These papers emphasized that not only the quantity of experts' knowledge was important, but that the *special organization of experts' knowledge* was at least equally important. The quantity-of-knowledge view received further critique because of the research methodology: Immediate memory (e.g., for chess boards) does not capture the essence of expertise in a domain. Another critique concerns the acquisition of knowledge: It is doubtful that knowledge is acquired piece by piece as a direct consequence of increasing experience.

#### *Organization-of-Knowledge View*

Consequently, the historically subsequent explanation for experts' superior abilities was based on the idea that experts not only possess more domain-specific knowledge, but also qualitatively different knowledge. In order to investigate this *organization-of-knowledge* view of

expertise (Sternberg, 1997) semantically rich (complex) tasks that afford ample prior knowledge were used and expert processes were often captured by think aloud protocols (Chi, 2006; Ericsson, 2005; Feltovich, Prietula, & Ericsson, 2006). Empirical results support this *organization-of-knowledge* view: Experts spend more time on *good* problem representations (Feltovich, Prietula, & Ericsson, 2006; van Gog, Pas, & van Merriënboer, 2005) which are derived from an almost automatic problem perception including fast identification of relevant features (Feltovich, Prietula, & Ericsson, 2006) and including a quasi-automatic recognition of solutions (Chase & Simon, 1973). Resultingly, experts build more abstract and deeper problem representations (Feltovich, Prietula, & Ericsson, 2006; Rikers & Paas, 2005), use qualitatively different strategies for problem solving (Simon & Simon, 1978, cited in Chi, Glaser, & Rees, 1982) and possess superior metacognitive abilities such as monitoring and planning (Gruber & Mandl, 1996; Ericsson, 2005; Feltovich, Prietula, & Ericsson, 2006; Sternberg, 1998; Veenman & Elshout, 1999; Zimmerman, 2006).

Consistent theoretical explanations were proposed: For example Ericsson and Kintsch (1995) developed the *long term working memory theory* (LTWM). LTWM assumes that experts possess extraordinary memory skills that allow them to utilize long term memory as working memory. Information that is prescribed an anticipated future utility is supposed to be encoded specifically in order to be accessible later on. The *deliberate practice theory* (Ericsson, Krampe, & Tesch-Römer, 1993) on the other hand, proposes that a qualitatively different learning process leads to expertise, deliberate practice: “The crucial factor [...] is the engagement in special practice activities that allow performers to improve specific aspects of their performance” (Ericsson, 2005, p. 237; also Charness, Tuffiash, Krampe, Reingold, & Vasyukoya, 2005; Feltovich, Prietula, & Ericsson, 2006). This suggestion is based on the empirical finding that after an initial increase expertise will reach a plateau at a moderate level if no special practice is employed (Feltovich, Prietula, & Ericsson, 2006). The notion of *knowledge encapsulation* (Schmidt & Boshuizen, 1993a) emphasizes the implicit and tacit nature of experts’ knowledge: “As a result of extensive practice [...] biomedical knowledge has become linked with, or encapsulated under, a limited number of clinically relevant concepts that have the same explanatory power as the elaborate biomedical structure” (Rikers, Schmidt, & Boshuizen, 2000, p. 152). It is based on empirical results (Boshuizen & Schmidt, 1992; Schmidt & Boshuizen, 1993b; Rikers, Schmidt, & Moulart, 2005) indicating that medical experts, contrary to intermediates and novices, do not overtly apply biomedical knowledge in clinical reasoning, although they possess more in-depth knowledge.

#### *Operationalisations of Expertise*

This selective review shows that experts with ample prior domain knowledge differ on many dimensions from novices with almost nonexistent prior domain knowledge: “Experts certainly know more, but they also know differently” (Feltovich, Prietula, & Ericsson, 2006, p. 21). These considerations also lead to conclusions with regard to the question how prior

domain knowledge can be operationalized adequately. A broadly accepted definition of an expert is a person who permanently excels in a specific domain (not randomly and singular; Gruber, 1994). In order to circumvent this strict definition, expertise can also be determined relatively on a continuous scale from novice to master (Bromme, 1992; Chi, 2006). From this perspective novices could also be compared with more advanced learners.

Within the expert paradigm, Chi (2006) defines a novice as “someone who is new – a probationary member. There has been some minimal exposure to the domain.” (p. 22). Considering the topic that will be used in the empirical studies of this thesis (“genetic fingerprinting”, chapter 3.2), it is easy to find novices with only minimal exposure to that topic. For example, students with humanities majors can be considered novices. However, it is almost impossible to find real experts – at least in adequate number. Thus, novices will be compared with apprentices instead of with experts. Within the expert paradigm Chi (2006) defines an apprentice as someone “who is learning – a student undergoing a program of instruction beyond the introductory level” (Chi, 2006, p. 22). Advanced students of biology can be considered apprentices. According to Rouet and colleagues (Rouet, Favart, Britt, & Perfetti, 1997) such advanced students can also be considered discipline experts because they know the tools of the discipline (e.g., in this case how to run a PCR or how to interpret an electrophoretogram). Consequently, the empirical studies of this thesis will use the classical method of the expert paradigm (expert-novice comparison) but with apprentices or discipline experts instead of “real” experts.

### 2.5.2 Empirical Results and Corresponding Theoretical Explanations

Empirical studies from as different research traditions as the traditional calibration paradigm, problem solving, text comprehension, hypertext learning, and hypertext representation consistently show a high impact of *prior domain knowledge* on the cognitive and meta-cognitive components of the learning process as well as on the learning outcome. This is especially true for material of at least moderate complexity.

#### *Results regarding the Preparatory Stages of Self-Regulated Learning*

Within the expert paradigm, Chi, Feltovich, and Glaser (1981) investigated novices' and experts' problem representations which closely resemble students' *task definitions* according to the COPES-model. In a first study, 8 experts and 8 novices categorized 24 physics problems according to similarities in solution. Quantitative results revealed no difference in number of applied categories or time needed to sort the problems. Qualitative analyses on the other hand indicated differences in applied sorting criteria: “experts categorize physics problems by the underlying physics principle [...], whereas novices categorize problems by the surface structure of the problems” (Chi, Feltovich, & Glaser, 1981, p. 134). These re-

sults were confirmed by a second study that explicitly crossed surface features with applicable physics laws. Chi, Glaser, and Rees (1982) report similar results: Although novices and experts did not differ significantly in time spent on problem representation they derived qualitatively different inferences from their task definition: “Novices are more likely either to generate the wrong inference or fail to generate the necessary inferences” (Chi, Glaser, & Rees, 1982, p. 40). Thus, experts seem to “perceive more in a problem statement” (Chi, Feltovich, & Glaser, 1981, p. 147; also Smith, 1990; Sternberg, 1998).

In an empirical study with a different focus (Chi, Glaser, & Rees, 1982), six novices and six experts rated the *difficulty* of 20 physics problems and circled key words or phrases that they used for their judgments. Results indicate that novices and experts used similar and a similar number of cues to infer task difficulty. Nonetheless, “experts, in general, were more accurate at judging the difficulty of problems than novices” (Chi, Glaser, & Rees, 1982, p. 68). Asking them why they selected the cue or what they inferred from the cue revealed that novices concentrated on the right cues, but inferred “nonphysics-related or nonproblem-related features” (Chi, Glaser, & Rees, 1982, p. 68). A study by Lodewyk and Winne (2005) sheds further light on the relation between expertise and judged task difficulty: Ninety-four 10<sup>th</sup> graders were confronted with one moderately well-structured and one ill-structured classroom task. Based on their performance in the previous year the students were classified as high, medium and low general academic achievers. Only the moderate achievers’ perceived task difficulty significantly varied between both tasks in the correct direction: 70 % of these students considered the ill-defined task more difficult than the well-defined tasks. Contrary, only 40 % of the high achievers considered the ill-defined task more difficult than the well-defined and only 58 % of the low achievers.

To summarize, these empirical studies consistently show little difference between novices and experts with regard to *quantitative* data: Experts and novices classify tasks into the same number of categories (Chi, Feltovich, & Glaser, 1981), they attend to the same and the same number of cues (Chi, Glaser, & Rees, 1982), and take the same time for task interpretation (Chi, Feltovich, & Glaser, 1981; Chi, Glaser, & Rees, 1982). *Qualitative* data on the other hand indicates that experts use different internal standards or criteria to evaluate tasks (Chi, Feltovich, & Glaser, 1981, Chi, Glaser, & Rees, 1982; Lodewyk & Winne, 2005; Smith, 1990) and draw different inferences from what they perceive (Chi, Glaser, & Rees, 1982). While most results indicate that experts develop more accurate task definitions, some results also point in the opposite direction: high achievers in Lodewyk and Winne’s (2005) study were worse at judging task difficulty than moderate achievers. Consider that these difference in the *preparatory* stages of the learning process might impact subsequent *enactment* stages: “Hence, the quality, completeness, and coherence of an internal representation [*task definition*] must necessarily determine the extent and accuracy of derived inferences, which in turn might determine the ease of arriving at a solution and its accuracy” (Chi, Glaser, & Rees, 1982, p. 30).



*Results regarding the Enactment Stages of Self-Regulated Learning*

Lind and Sandmann (2003) demonstrated the significant impact of expertise on the use of learning strategies: Twenty students who participated in the “biology Olympics” were considered experts in biology. Twenty students who participated in the “physics Olympics” were considered experts in physics. All students read introductory texts of human genetics (biology) and space navigation (physics) and solved 5 problems within each domain. They thought aloud during task solution. Results indicate that more prior domain knowledge was positively correlated with the use of deep elaboration strategies (e.g., link presented information with prior knowledge, infer a situational model of the presented problem), while it was negatively correlated with the use of surface strategies (e.g., re-read a text, infer a text-based model of the presented problem). Consistent results were found for adults by Lawless, Brown, Mills, and Mayall (2003): Thirty-four undergraduates learned with a hypertext about Lyme disease. Domain knowledge was significantly and positively correlated with structured recall and unstructured recall (learning outcome). Furthermore, logfiles indicate that learners with more domain knowledge used more reading strategies and “selected and sequenced information more efficiently” (Lawless, Brown, Mills, & Mayall, 2003, p. 925). A study by Ford and Chen (2000) yielded similar results.

In contrast to these consistently positive results, Rouet and colleagues (Rouet, 2003; Rouet, Favart, Britt, & Perfetti, 1997) found almost no impact of prior domain knowledge on study strategies. Rouet, Favart, Britt, and Perfetti (1997), for example, investigated the impact of domain expertise on students’ reading, document evaluation and use of multiple documents about a historical controversy. Results indicate that employed study strategies and students’ evaluations with regard to documents’ trustworthiness did not differ across groups with different domain expertise (11 psychology graduates, 8 history graduates). Nonetheless, students significantly varied with regard to the evaluation of documents’ usefulness: Novices evaluated mainly with regard to content, specialists on the other hand applied multiple criteria that focused more on interpretations and evidence and varied across document types demonstrating some flexibility. Furthermore, novices and specialists differed significantly in their essays: “Most novice students expressed an opinion about which side was right, whereas the specialists expressed an opinion about the structure of the problem space” (Rouet, Favart, Britt, & Perfetti, 1997). As an explanation for the lack of effects of prior domain knowledge on study strategies Rouet (2003) proposed that “these aspects of the search activity may call upon general cognitive resources that are brought to bear in similar ways regardless of the domain” (p. 423).

To summarize, these empirical studies consistently show the following effects: First, experts achieve better *learning outcome* than novices (Ford & Chen, 2000; Lawless, Brown, Mills, & Mayall, 2003; Lind & Sandmann, 2003; Rouet, 2003; Rouet, Favart, Britt, & Perfetti, 1997). Second, results for the impact of prior domain knowledge on *learning processes*

are mixed: Some empirical studies indicate that learners with more prior domain knowledge apply more elaborate and more efficient strategies (Ford & Chen, 2000; Lawless, Brown, Mills, & Mayall, 2003; Lind & Sandmann, 2003), while others found almost no effects (Rouet, 2003; Rouet, Favart, Britt, & Perfetti, 1997).

#### *Results Indicating Interaction with Complexity*

The apparent inconsistencies between the results mentioned in the last section may be resolved by taking into account *task complexity* or *(hyper)text complexity*. Although this was not their prime goal, many empirical studies address this issue implicitly when studying the interaction between prior domain knowledge and hypertext structure (Calisir & Gurel, 2003; McDonald & Stevenson, 1998; Möller & Müller-Kalthoff, 2000; Potelle & Rouet, 2003) or the interaction between prior domain knowledge and text coherence (McNamara, Kintsch, Songer, & Kintsch, 1996; Salmerón, Kintsch, & Canas, 2006).

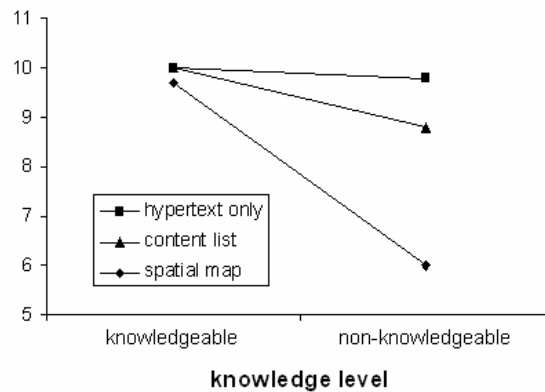


Figure 2.5-1: Mean number of correctly answered questions (Y-axis) as a function of prior knowledge and navigational aid (Figure adapted from McDonald & Stevenson, 1998, p. 135)

With regard to the hypertext structure, the presence or absence as well as the specific design of an instructional aid such as a content representation can make a whole hypertext simple by preventing students' disorientation or very demanding by forcing students to infer a structure on their own. Thus, such manipulations can be interpreted as variations in *(hyper)text complexity*. McDonald and Stevenson (1998), for example, asked 36 students to read a 45-node hypertext on human learning processes and subsequently to answer 10 questions with the help of the hypertext. Students who differed in prior domain knowledge (low vs. high) received (1) no navigational aid, (2) a content list or (3) a map of the hypertext. The hypertext without aid can be considered more complex than the hypertext with the content list which in turn might be more complex than the hypertext with the map. Results indicate main effects for prior knowledge and navigational aid: Consistently, more prior domain knowledge and the simpler hypertext versions lead to better navigation and

performance. Nonetheless, also interaction effects were found: For example, knowledgeable students were able to answer almost all questions correctly, independent of navigational aid, whereas low-knowledge students only were able to do so in the simplest map version of the hypertext (Figure 2.5-1). Consistent results were presented by Potelle and Rouet (2003), by Calsir and Gurel (2003) and by Möller and Müller-Kalthoff (2000).

With regard to text coherence, Kintsch and colleagues found similar results within the text comprehension paradigm (McNamara, Kintsch, Songer, & Kintsch, 1996; Salmerón, Kintsch, & Canas, 2006). McNamara, Kintsch, Songer, and Kintsch (1996) demonstrated that students with low prior domain knowledge benefit from a maximally coherent text, while students with high prior knowledge benefit from a minimally coherent text. These results were replicated with hypertexts by Salmerón, Kintsch, and Canas (2006). In their first experiment, the authors confronted 71 undergraduates of high and low prior domain knowledge with a hypertext about atmospheric pollution. Each node was linked to the most coherent and the least coherent node. Based on their link-selection, students were post hoc assigned to three navigation strategy groups: (1) first mentioned (most often selected the top link), (2) interest (asked post-hoc), and (3) coherence (most often selected the most coherent link). Results indicate that for inference questions an interaction effect between prior domain knowledge and strategy use was detected: No difference between employed strategies was found for high knowledge students, while the low knowledge students who utilized the coherence strategy outperformed low knowledge students using other strategies. These results were replicated with a different methodology in a second experiment: Salmerón, Kintsch, and Canas (2006) had 152 undergraduates read the same content as in experiment 1 in one of four instructed conditions: (1) read the hypertext according to interest, (2) read the hypertext according to coherence, (3) read a linear hypertext that was linked coherently, (4) read a linear hypertext that was linked incoherently. Thus, a two (interest vs. coherence) by two (instruction induced vs. text induced) by two (high knowledge vs. low knowledge) design was realized. Results indicate that with regard to inference questions, the results of the linear hypertexts (control condition, Figure 2.5-2) replicate McNamara, Kintsch, Songer, and Kintsch's (1996) results: "low-knowledge readers tended to learn more with the high-coherence than with the low-coherence order [...], whereas intermediate-knowledge readers showed the opposite pattern" (Salmerón, Kintsch, & Canas, 2006, p. 1165). The results of the instructed strategies (strategic condition, Figure 2.5-2) replicate the results of experiment 1 for the inference questions: "Low-knowledge participants in the coherence strategy condition outperformed those in the interest strategy condition [...], but this was not the case for the intermediate-knowledge participants" (Salmerón, Kintsch, & Canas, 2006, p. 1165). Thus, these studies indicate that prior domain knowledge interacts with *text complexity* as well as with *task complexity*: consistent effects were only found with tasks that are more complex (inference task) and with texts that are more complex (incoherent texts).

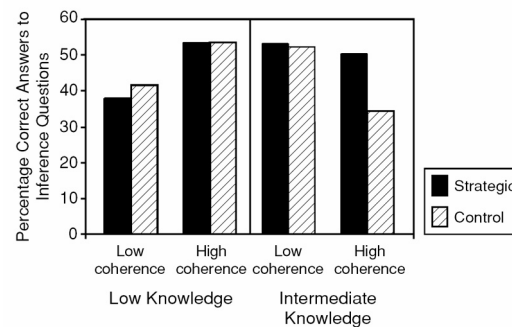


Figure 2.5-2: Percentage of correct answers as a function of type of task, prior knowledge, and text coherence (Figure taken from Salmerón, Kintsch, & Canas, 2006, p. 1166)

To summarize, these empirical studies consistently show that prior domain knowledge interacts with *complexity*: The superiority of experts for complex content disappears when the instructional material is simplified by adding instructional guidance (Calsir & Gurel, 2003; Kalyuga, Chandler, & Sweller, 2001; McDonalds & Stevenson, 1998; Möller & Müller-Kalthoff, 2000; Potelle & Rouet, 2003; Salmerón, Kintsch, & Canas, 2006; Tuovinen & Sweller, 1999); in some cases, instructional aids are even detrimental for experts while they are beneficial for novices (Kalyuga, Chandler, & Sweller, 1998; McNamara, Kintsch, Songer, & Kintsch, 1996; Yeung, Jen, & Sweller, 1998).

#### *Potential Theoretical Explanations*

In this section two potential explanations will be discussed for the empirical finding that high knowledge students perform better with highly complex texts and worse with simple texts (see previous sections on empirical results regarding the enactment stage).

Kintsch and colleagues (e.g., McNamara, Kintsch, Songer, & Kintsch, 1996) claim that in high knowledge students coherent texts elicit automated superficial processing whereas incoherent texts elicit active deep processing which in turn might lead to a more adequate situation model of the texts' content ("active processing" hypothesis). Support for this hypothesis comes from the comparison between text-induced and instruction-induced coherence (Salmerón, Kintsch, & Canas, 2006). It can be assumed that instruction induced strategy use requires more active processing to achieve coherence than receiving an already coherent text. Consistently, results indicate a significant advantage of "active processing" (Figure 2.5-2): "The intermediate-knowledge participants following the coherence strategy learned more than those performing linear reading of a high-coherence ordered text without link selection" (Salmerón, Kintsch, & Canas, 2006, p. 1166).

Another explanation was offered by cognitive load theory (Sweller, van Merriënboer, & Paas, 1998) as explanation of the expertise reversal effect which is conceptually similar to the pattern of results detected for text comprehension: Low knowledge students benefit

from additional instructional aids (e.g., integrated diagrams to avoid the split-attention effect), while high knowledge students learn equally well independent of provided guidance; in some instances additional guidance is even detrimental for high knowledge learners (Kalyuga, Chandler, and Sweller; 1998; Tuovinen & Sweller, 1999; Yeung, Jin, and Sweller; 1998). Kalyuga, Ayres, Chandler, and Sweller (2003) argue that learners with high prior domain knowledge “bring their activated schemas to the process of constructing mental representations” (Kalyuga, Ayres, Chandler, & Sweller, 2003, p. 24). Thus, additional instructional guidance is superfluous. If it is provided nonetheless, it is redundant with students’ own schemas and “will require additional working memory resources and might cause a cognitive overload” (Kalyuga, Ayres, Chandler, & Sweller, 2003, p. 24). This explanation is supported by empirical results indicating “that experienced learners studying a minimal format reported lower estimates of mental load compared to formats with redundant information” (Kalyuga, Ayres, Chandler, & Sweller, 2003, p. 26).

With regard to these offered explanations no final conclusion can be drawn because so far both proposed explanations are equally supported by empirical evidence.

#### *Empirical Results regarding Metacognitive Processes*

Empirical studies focusing on concurrent learning processes consistently demonstrate a superiority of experts in their metacognitive processes. For example, Veenman and Elshout (1999) investigated how novices and experts of low and high intellectual ability solve thermodynamics problems. Results indicate that the level of metacognitive skillfulness was significantly higher for the experts – independent of intellectual ability. Priest and Lindsay (1992) scrutinized novices’ and experts’ problem-solving strategies for physics problems. The most significant difference was that “experts were more likely to be able to plan their solutions at a descriptive meta-level than novices” (p. 389).

With regard to the classical *calibration* paradigm tapping metacognitive monitoring, results of empirical studies are mixed: Nietfeld and Schraw (2002) tested three different hypotheses about the relationship between expertise and calibration: (1) The debilitating hypothesis assumes that “prior knowledge is negatively related to monitoring accuracy because individuals become increasingly overconfident” (Nietfeld & Schraw, 2002, p. 133), (2) the no-impact hypothesis proposes that prior knowledge is independent from metacognitive monitoring processes (Commander & Stanwyck, 1997; Koriat, 1994; Morris, 1990; all cited in Nietfeld & Schraw, 2002), and (3) the facilitative hypothesis states that prior knowledge improves metacognitive accuracy (Stone, 2000). For the latter hypothesis two reasons are feasible: “First, prior knowledge provides a conceptual basis for evaluating one’s performance [...]. Second, experts are more apt to be automated problem solvers, which should yield more cognitive resources for monitoring one’s performance.” (Nietfeld & Schraw, 2002, p. 133). Ninety-three undergraduates solved parts of the Raven Matrices as proxy of general ability, completed 24 multiple-choice items on probability and gave confidence

ratings after each item. According to their previous coursework in math and statistics, students were classified in three sub-samples of low-, mid-, and high-knowledge. Results indicate that high-knowledge students achieved a higher score on the multiple-choice questions and also were more accurate in their confidence judgments than the students of the other two sub-samples. No effect for students' confidence ratings and a bias score were found. Nietfeld and Schraw (2002) concluded that these empirical findings mostly support the facilitative hypothesis but also lend some support to the no-impact hypothesis. The results of Pressley and Ghatala (1988) are consistent with these conclusions. On the other hand the results of Glenberg and Epstein (1987) clearly support the debilitating hypothesis as in their study expertise in a domain was "inversely related to calibration" (p. 84). However, their proposed explanation ("self-classification hypothesis") was refuted by an empirical study of Maki and Serra (1992; Lin, Zabrocky, & Moore, 1997; Weaver, 1990).

To summarize, empirical studies demonstrated a significant advantage of experts for concurrently enacted metacognitive strategies (Priest & Lindsay, 1992; Schneider, Schlagmüller & Visé, 1998; Veenman & Elshout, 1999). Results within the traditional *calibration* paradigm on the other hand are mixed, but empirical evidence slightly favors the facilitative or no-impact hypotheses over the debilitating hypothesis. Furthermore, the effect of prior domain knowledge on metacognitive calibration might be dependent on the given task. Stone (2000), for example, concluded that "expertise appears to effect better calibration when the task is stable [...], or when the inferences lead to a smaller number of possible solutions [for simple tasks]" (p. 449).

### 2.5.3 Conclusion

To investigate the impact of prior domain knowledge on metacognitive calibration processes, this internal condition needs to be validly operationalised to constitute an adequate independent predictor variable. Based on a selective review of the expert paradigm the methodology of expert-novice comparison was selected. Thus, students with different levels of prior domain knowledge will be selectively recruited for the empirical studies of this thesis: (1) Students with *humanities majors* can be considered "novices" while (2) *advanced students of biology* can be considered "discipline experts" regarding the topic of "genetic fingerprinting". These differences in prior domain knowledge will be validated by a knowledge test tapping the quantity of learners' molecular biology knowledge. Consequently, the impact of prior domain knowledge can be diagnosed statistically by comparing these two quasi-experimental groups (between-subject).

Based on the reviewed empirical findings and theories the following effects will be expected in the empirical studies of this thesis: First, experts will excel with regard to most aspects of the self-regulated learning process and with regard to the learning outcome, for

example they will use more elaborate deep criteria to evaluate tasks, they will draw better inferences from what they perceive, and their task definitions will be more accurate (*preparatory* stage); they will enact more elaborate and efficient domain-specific strategies and they will have better learning outcomes (*enactment* stage), and they will demonstrate more proficient metacognitive skills (*metacognitive processes*). Second, domain expertise will not be equally beneficial for learning processes and outcomes in all scenarios, instead an interaction with complexity can be expected: At least for complex tasks (Salmerón, Kintsch, & Canas, 2006) and for more complex (hyper)texts (McDonalds & Stevenson, 1998) prior domain knowledge will result in better learning processes and outcomes. For very simple materials no effects or even reversed effects might be detected. Third, no consistent superiority of experts can be expected with regard to quantitative measures in the *preparatory* stages, with regard to learning process variables in the *enactment* stages of self-regulated learning, and with regard to *calibration* in the traditional sense.

Note that the COPES-model (Winne & Hadwin, 1998; chapter 2.1) would predict slightly different effects: First, it would predict that experts will not openly metacognitively regulate their learning process because “the more extensive one’s domain knowledge is, the less is the need to search for, use, and metacognitively regulate tactics or strategies” (Winne, 1996; p. 333). Note, that this is a significant difference to the predictions based on other empirical studies (see above). Second, Winne and colleagues indicate that domain knowledge might interact with complexity: “In this circumstance [of complexity], experts’ performance varies in relation to how validly they are able to perceive the task they face and the validity of that perception can depend on whether metacognitive monitoring succeeds” (Winne, 1996; p. 333). This might imply that for complex tasks the preparatory stages of learning (*task definition*) might become more important. Third, the issue of the impact of prior domain knowledge on learning outcome is not addressed directly. However, based on the fact that prior domain knowledge is assumed to be beneficial for learning processes, it would be safe to predict that it should also be beneficial for learning outcomes.

Thus, prior domain knowledge might either have a beneficial (see empirical studies) or detrimental (see COPES-model) *main* effect on students’ overt metacognitive self-regulated learning *process*. Instead or in addition to such a main effect, it might *interact* with complexity (e.g., prior domain knowledge might be more important for complex tasks or texts). And all predictions concur that knowledge should be beneficial for *learning outcome* (*main effect*).

## 2.6 Epistemological Beliefs – An Important Internal Condition

Epistemological beliefs (learners’ beliefs about the nature of knowledge and knowing) are one of the potentially most influential learner characteristics and should impact the whole self-regulated learning process. The inclusion of this variable in models of self-regulated

learning is rare but advocated: “Epistemological beliefs should be included in models of self-regulated learning” (Bråten & Strømsø, 1995, p. 559; also Schommer-Aikins, 2004). The COPES-model (Winne & Hadwin, 1998; chapter 2.1) complies with this demand by specifying a functional relationship between epistemological beliefs and all components of self-regulated learning. Winne and Hadwin (1998) propose that epistemological beliefs influences the whole self-regulated learning by impacting learners’ internal standards: “IF a student believes that learning is easy or that effort (a cue) is unimportant in performance, THEN that student may choose not to adapt learning tactics to improve learning outcomes” (Butler & Winne, 1995, p. 254). Consequently, “epistemological beliefs about studying would be predicted to influence learners’ metacognitive studying strategies.” (Winne, 1995, p. 179). Additionally, the importance of epistemological beliefs is obvious in everyday learning experiences: Consider a student who believes that knowledge is simple and certain (“naïve”) in contrast to a student who believes that knowledge is complex and uncertain (“sophisticated”). The first student probably would employ more superficial learning strategies like memorizing while the second student probably would employ strategies of deeper elaboration to compensate for perceived complexity and would consult multiple information sources to compensate for uncertainty.

Research on epistemological beliefs is quite young as a field of psychology. Thus, it is not surprising that diverse theories of epistemological beliefs are proposed which still raise many open questions and controversies. These theoretical conceptualizations of epistemological beliefs will be reviewed first and potential operationalisations will be considered. Empirical results reviewed subsequently demonstrate significant impact of epistemological beliefs on most aspects of learners’ self-regulated learning process.

### 2.6.1 Theoretical Conceptualizations and Potential Operationalisations

Epistemology or the study of knowledge and knowing has long been one of the cornerstones of philosophy (Buehl & Alexander, 2001). The term is derived from the Greek *episteme* (knowledge) and *logos* (explanation). Thus, epistemology involves questions pertaining to the origin, nature, form, limits and methods of human knowledge and questions about the processes by which such knowledge is verified and justified (Hofer, 2002). Philosophical pragmatists first raised concerns about the relationship between knowledge and schooling. As a logical consequence, epistemology was no longer only abstractly debated by philosophers, but epistemological conceptions of teachers, students and laypersons were assessed qualitatively and quantitatively. The psychological research tradition of epistemo-



logical beliefs<sup>5</sup> was born (for recent overviews see Buehl & Alexander, 2001; Duell & Schommer-Aikins, 2001; Hofer & Pintrich, 1997, 2002).

Theoretically, epistemological beliefs are often conceptualized as structurally related to metacognitive processes (Bendixen & Hartley, 2003; Hofer, 2001, 2004; Kitchener, 1983). Sometimes epistemological beliefs are discussed as an abstract level in the cognitive system that might be located above a metacognitive level (Kitchener, 1983), sometimes epistemological beliefs are conceptualized as part of a learners' metacognitive knowledge (Hofer, 2004; Kuhn, 2000). Kitchener (1983), for example, proposed three levels of cognitive processing to account for the complex monitoring that is involved in dealing with ill-structured problems: (1) *cognition* (all cognitive operations such as reading, memorizing, perceiving, computing), (2) *metacognition* (all cognitions that have cognitive operations as their subjects, for example monitoring of the cognitive operations), and (3) *epistemic cognition* (cognitions about the limits of knowing, the certainty of knowledge, or the criteria for knowledge). Only the latter kind of cognitions is assumed to be involved in monitoring the epistemic nature of problems and the evaluation of solutions. To give another example, Hofer (2004) details how epistemological belief dimensions can be matched to components of models of metacognition. She argues that beliefs about the nature of knowledge (*certainty of knowledge* and *simplicity of knowledge*) share similarities with declarative metacognitive knowledge (Hasselhorn, 1998; Schraw & Moshman, 1995). Beliefs about the nature of knowing on the other hand (*source* and *justification of knowledge*) can be matched to the procedural component of metacognition, to metacognitive monitoring and control (Hasselhorn, 1998; Schraw & Moshman, 1995). These models concerned with structural aspects of epistemological beliefs (i.e., where they are located in the cognitive system) are promising but far from complete. Most theories and empirical studies in the field of epistemological beliefs so far have been inspired by the seminal works of Perry (1970) who mostly motivated research in the *developmental* psychology tradition and by the works of Schommer (1990) who mostly stimulated research in the *educational* psychology tradition.

#### *Research inspired by Perry (1970) – The Developmental Psychology Tradition*

Since Perry (1970, also Bromme, 2005; Buehl & Alexander, 2001; Hofer & Pintrich, 1997; Moore, 2002), the interest of *developmental* psychologists in changes which occur in individuals' beliefs over time prevails. Within the stage-models dominant in this research tradition epistemological beliefs are conceptualized as domain-general and relatively "trait-like" ca-

---

<sup>5</sup> I will consistently utilize the term "epistemological beliefs" throughout this thesis in order to characterize individual's beliefs about the nature of knowledge and knowing (Hofer & Pintrich, 1997) even though other terms have been proposed in the literature (e.g., Hofer, 2002; Niessen, Vermunt, Abma, Widdershoven, & van der Vleuten, 2004): personal epistemology, epistemological theories, epistemological posture, epistemological reflection, epistemological stance, epistemological resources, epistemological repertoires, epistemological thinking, epistemological orientation, ways of knowing, epistemic cognition, epistemic reflection.

pability of individuals to deal with controversial issues. Interviews and open-ended questions are the favored research methodologies of this strand of research.

To understand the intellectual development of college students, Perry himself interviewed male students over the course of their undergraduate education. He detected a pattern which can be summarized as follows: (1) freshmen often adopt a *dualistic* position in which knowledge is seen as either right or wrong. (2) More advanced students on the other hand often express *multiplistic* views that acknowledge diversity and uncertainty but still emphasize that truth is attainable. (3) *Contextual relativism* on the other hand encompasses the acknowledgement that knowing always contains a personal perspective; thus, all knowledge is considered relative. (4) The most advanced position addresses *commitment within relativism* (i.e., although students acknowledge the relativism of knowledge they nonetheless commit to one point of view). Although Perry never set out to investigate epistemological beliefs, his research inspired other researchers (Baxter Magolda, 1992; Belenky, Clinchy, Goldberger, & Tarule, 1986; King & Kitchener, 1994; Kuhn, 1991).

For example, Belenky, Clinchy, Goldberger, and Tarule (1986; also Clinchy, 2002), criticized Perry's focus on an exclusively male college level sample and thus alternatively interviewed *women* ( $n = 135$ ) and deduced five ways of women's knowing that partly differed from Perry's scheme: (1) *silence* (not fit to contribute), (2) *received knowledge* (knowledge is either right or wrong), (3) *subjectivism* (oneself is the only valid source of knowledge), (4) *procedural knowing* (knowing requires active work; multiple interpretations are possible, but some interpretations are more valid), and (5) *constructed knowing* (knowledge is complex and ambiguous, integration of these "many voices" is tried).

Kuhn (1991) *expanded the age range* by presenting individuals from their teens till their 60ies with ill-structured problems and asking them to state and to justify their positions. She uncovered three epistemological views in participants' responses: (1) *absolutist* (knowledge is certain and absolute), (2) *multiplist* (knowledge is uncertain; all views are equally valid) and (3) *evaluativist* (knowledge is uncertain, but views can be compared with regard to their validity) and argued that the underlying developmental task is the "coordination of the *subjective* and *objective* dimensions of knowing" (Kuhn & Weinstock, 2002, p. 123). With a standardized and abbreviated interview technique Weinstock (Kuhn & Weinstock, 2002) demonstrated that these kinds of epistemological views matter in the real world: Jurors' verdicts and argumentations were related to their epistemological beliefs.

Baxter Magolda (1992, 2002) conducted an *additional longitudinal study* where she traced 18 to 30 year olds' views about education and knowledge and developed the 4-stage Epistemological Reflection model: (1) *absolute knowing* (knowledge is certain and handed out by authorities), (2) *transitional knowledge* (some pieces of knowledge are uncertain, some pieces are certain), (3) *independent knowing* (most knowledge is uncertain), and (4) *contextual knowing* (knowledge is uncertain, not all solutions are equally valid).

King and Kitchener (1994, 2002) investigated *argumentation* by interviewing more than 1700 individuals with the help of ill-structured problems and follow-up questions (reflective judgment interview, RJI) and “found that individuals’ assumptions and beliefs about knowledge were related to how they chose to justify their beliefs.” (Buehl & Alexander, 2001, p. 392). Their Reflective Judgment Model (RJM) entails three main positions (Kitchener, 1983): (1) *prereflective* (knowledge is certain, correct and is justified by authorities and first hand experience), (2) *quasi-reflective* (knowledge contains uncertainties, but these are attributed to missing observations or missing methodology; knowledge is justified by evidence but without full understanding), and (3) *reflective* (knowledge is uncertain, but an active evaluation is required by choosing the more reasonable alternative).

To summarize, a vast amount of qualitative data – longitudinal (Baxter Magolda, 1992; Perry, 1970) and cross-sectional (King & Kitchener, 1994; Kuhn, 1991; Kuhn & Weinstock, 2002) in nature – supports a general developmental sequence from epistemological views of *absolutism* (knowledge is either right or wrong), followed by *relativism* (everything is relative and subjective) and ending at some sort of *evaluativism* (although everything is relative, some views have better empirical support). However, open questions and critical issues currently remain within this research tradition.

#### *What Triggers Epistemological Change?*

It is still unclear which exact mechanism governs the change of epistemological beliefs, for example from *absolutist* to *relativist* epistemological views. As a first theoretical approximation Bendixen (2002) suggested a process model of epistemic beliefs change based on Piaget’s (1985, cited in Bendixen, 2002) ideas of disequilibrium as trigger of cognitive development and based on the notion of conceptual change if old structures no longer work satisfactory. In this model epistemic belief change is assumed to be triggered by the exposure to differences, independence or beliefs not matching up with experiences and students are supposed to experience epistemic doubt as confusion connected with scary feelings. An in-depth interview study with 12 participants lends some empirical support to this notion: The majority of students resolved epistemic doubt by taking control (personal reflection or seeking additional information) and consequently reported developing new and better beliefs. A minority of students, however, resolved epistemic doubt by surrendering control (e.g., by relying on a higher power such as God) and consequently reported that their initial beliefs were strengthened and reaffirmed.

#### *Do Epistemological Beliefs develop toward more “Sophisticated” Beliefs?*

The developmental sequence as well as the corresponding terminology is discussed critically. The proposed developmental sequence (from *absolutist*, to *relativist*, and to *evaluativist*) indirectly implies terminology: The term “naïve” might characterize an *absolutist* view (that knowledge is absolute, certain, an accumulation of facts and can be transferred effectively

by an authority). The term “sophisticated” on the other hand in this context might refer to an *evaluativistic* view (that individuals become aware that knowledge is a complex network, relative and uncertain, and contextual thus requiring an active evaluation of different viewpoints with regard to validity).

The first point of criticism concerns the fact that almost the same distribution of *absolutists*, *relativists* and *evaluativists* has been empirically found at different age levels (Chandler, Hallett, & Sokol, 2002; Kuhn & Weinstock, 2002): For example, studies within the theory of mind paradigm indicate that even young children are able to acknowledge uncertainties and evaluate evidence adequately in order to evaluate knowledge claims (“early-onset” view). This is clearly at odds with Perry’s initial conclusion that most students arrive at college with “naïve” *absolutistic* epistemological beliefs and only the brightest reach the *evaluativist* level (“late-onset” view). One potential explanation for this empirical phenomenon is that individuals’ development of epistemological beliefs is recursive, but possesses different quality at different age or educational levels: “new and improved understandings of representational diversity occur, and then reoccur, first at what will be marketed as a ‘retail’, and then a ‘wholesale’ level” (Chandler, Hallett, & Sokol, 2002, p. 147).

A second point of criticism concerns the value-laden adjectives “naïve” and “sophisticated”. These adjectives implicitly imply that more “naïve” epistemological beliefs can be classified as some kind of “misbeliefs” (Hammer & Elby, 2002). In theory, this kind of evaluation requires an omniscient outside view and clearly no human possesses such an outside view! However, researchers such as Schommer-Aiken (2002) argue adamantly for assigning such appraising labels in order not to “drain in the pit of relativism”. This is in line with the fact that philosophers, historians and sociologists agree upon a view of knowledge that is more valid than others views, for example that knowledge is tentative, partially subjective, relies on an empirical basis and so on (Sandoval, 2005).

A third point of criticism concerns the fact that – at least implicitly - “sophistication” is considered to be context-independent. This might be unrealistic because “it is hardly sophisticated, for example, to consider it ‘tentative’ that the earth is round” (Hammer & Elby, 2002, p. 186). Philosophers and scientists concur and conclude that one property of “knowledge” might be that “some claims are more tentative than others” (Sandoval, 2005, p. 641). Bromme (2005) further argues that it might be “sophisticated” to recognize the need for a division of cognitive labor (i.e., be able to flexibly judge which situations require yielding to expert authorities). Therefore, sensitivity to contexts might constitute most “sophisticated” epistemologies, that is having dynamic and flexible epistemological beliefs (also Leach, Hind, & Ryder, 2003).

#### *Research inspired by Schommer (1990) – The Educational Psychology Tradition*

Since Schommer (e.g., 1990, 1994, 2002), research in *educational* psychology focuses on a multidimensional conceptualization of epistemological beliefs and on the impact of these

beliefs on learning. Schommer was the first to dispute the uni-dimensional conception of epistemological beliefs held by the developmental psychology research tradition (see previous sections) and instead suggested that “individuals’ epistemological beliefs are a system of beliefs composed of more or less separate dimensions” (Buehl & Alexander, 2001, p. 395). Furthermore, within this educational psychology research tradition epistemological beliefs are considered less “trait-like” and therefore changeable by instructional interventions. Multidimensional questionnaires are the favored research methodology of this strand of research because they facilitate economic data collection even in large samples.

Schommer herself developed the 63-item SEQ (Schommer Epistemological Questionnaire) where subjects had to indicate their agreement or disagreement to statements such as “The only thing that is certain is uncertainty itself.” Based on the synthesis of previous research, she proposed five dimensions. Empirical data corroborated four of these dimensions (Schommer, 1990, 1993; Schommer, Crouse, & Rhodes, 1992): (1) *innate ability* (ability is fixed vs. live-long improvement), (2) *simple knowledge* (knowledge consists of isolated bits vs. integrated concepts), (3) *quick learning* (knowledge acquisition is quick vs. gradual) and (4) *certain knowledge* (knowledge is tentative vs. unchanging). Even more importantly, Schommer also detected some links between epistemological beliefs and learning processes: For example, the belief in *simple knowledge* was associated with the use of study strategies and text comprehension (Schommer, Crouse, & Rhodes, 1992). Nonetheless, Schommer’s seminal work also raised a number of ongoing controversies.

#### *What are the Dimensions of Epistemological Beliefs?*

The first controversially discussed issue concerns the number and kind of dimensions of the construct epistemological beliefs (e.g., Bromme, 2005; Bromme & Stahl, 2003). Not all researchers detected empirical evidence for the multidimensionality of epistemological beliefs (Rozendaal, de Brabander, & Minnaert, 2001). However, most researchers found such indicators and thus favored a Schommer-inspired multidimensional conceptualization of epistemological beliefs (Bråten & Strømsø, 2005; Buehl, Alexander, & Murphy, 2002; Sandoval & Morris, 2003). But even if multidimensionality per se is accepted, some problems remain: On the one hand, empirical data did not consistently support the dimensional structure proposed by Schommer (Bråten & Strømsø, 2004, 2005, 2006; Cano, 2005; Clarebout, Elen, Luyten & Bamps, 2001; Hofer, 2000; Kardash & Howell, 2000; Qian & Alverman, 1995; Schraw, Bendixen & Dunkle, 2002; Sinatra, Southerland, McConaughy, & Demastes, 2003; for a more general criticism of the use of factor analysis in research on epistemological beliefs see Wood & Kardash, 2002). On the other hand, her dimensions can be criticized theoretically. For example, Hofer and Pintrich (1997) argued that the dimensions *innate ability* and *quick learning* are not “epistemological” in a strict sense, but rather can be considered students’ beliefs about intelligence or learning. This critique is legitimate and is not seriously denied by Schommer (Schommer-Aikins, 2002). Thus, Hofer

and Pintrich (1997) proposed an alternative four-dimensional framework of epistemological beliefs with the dimensions: *certainty of knowledge*, *simplicity of knowledge*, *source of knowledge*, and *justification for knowing*. For an even more rigorous theoretical evaluation of epistemological belief dimensions see Rozendaal, de Brabander, and Minneart (2001): Based on the core criterion of “validity of knowledge” they only consider *certainty of knowledge*, *source of knowledge* and *justification of knowledge* as valid epistemological dimensions, but reject the dimension of *simplicity of knowledge* as well as all beliefs about learning. Due to these discussions Pintrich (2002) concluded that there is more than one dimension, but less than ten.

#### *Are Epistemological Beliefs Domain-General or Domain-Specific?*

A second open issue concerns the domain-general or domain-specificity of epistemological beliefs (e.g., Bromme, 2005; Bromme & Stahl, 2003). Learners may have general epistemological beliefs, as well as beliefs that differ substantially between academic domains (Buehl, Alexander & Murphy, 2002; de Corte, Op’t Eynde & Verschaffel, 2002; Estes, Chandler, Horvath & Backus, 2003; Kardash & Scholes, 1996; Schommer & Walker, 1995; Trautwein, Lüdtke, & Beyer, 2004). But up to now, it is unclear how such different levels might interact with each other (Hofer, 2000).

Empirical investigations of this issue shed some light: Jeheng, Johnson, and Anderson (1993), for example, compared students majoring in so-called ‘soft’ fields (social science, arts, humanities) with students majoring in so-called ‘hard’ fields (business, engineering) in a between-subject design. They administered a modified version of the SEQ to 386 students and found that “students from the ‘soft’ fields believed less in the *certainty of knowledge*, relied more on their own reasoning abilities, and were less likely to view learning as an orderly process than students majoring in ‘hard’ fields” (Buehl & Alexander, 2001, p. 403). Paulsen and Wells (1998) found similar results.

Within-subjects comparisons on the other hand reveal mixed results: Stodolsky, Salk, and Glaessner (1991), for example, interviewed 60 5<sup>th</sup>-graders about their attitudes, perceptions, and dispositions towards mathematics and social studies. Children’s responses revealed that “students not only defined mathematics and social studies differently, but they also relied on rather distinct sources and cognitive processes to acquire relevant knowledge” (Buehl & Alexander, 2001, p. 405). To give another example using questionnaires, Hofer (2000) used her domain-specific questionnaire DEBQ (discipline focused epistemological beliefs questionnaire) and had 326 students complete it twice, once for psychology and once for science. A series of *t*-tests revealed, for example, that students considered science knowledge to be more certain and unchanging (*certainty / simplicity of knowledge*) and less justified by personal experience (*justification of knowing*) than psychology knowledge. Nonetheless, Hofer (2000) also found significant correlations between the dimensions of both domains. Further results obtained with similar methodology contribute to both sides of the argument: Results of Buehl, Alexander, and Murphy (2002) lend further support to do-

main-specific beliefs while results from Schommer and Walker (1995) stronger support a predominantly domain-general view of epistemological beliefs.

Thus, these mixed results demonstrate that even though students seem to perceive different academic domains differently, some kind of general epistemological beliefs system can also be assumed. To complicate this issue, there are even some hints that learners' epistemological beliefs might not only vary as a function of domain, but also as a function of context (Bråten & Strømsø, 2005; Bromme, 2005; Elby & Hammer 2001; Hammer & Elby, 2002; Pintrich, 2002; Sandoval, 2005; Stadtler & Bromme, 2005).

#### *Excursus – Measurement of Epistemological Beliefs*

The conceptual controversies outlined above are also closely linked to methodological problems. Epistemological beliefs are commonly measured via interviews and questionnaires (for an overview of instruments see Duell & Schommer-Aikins, 2001; Niessen, Vermunt, Abma, Widdershoven, & van der Vleuten, 2004). Research within the *developmental* psychology tradition mostly utilizes lengthy interviews where subjects are confronted with controversial evidence or decision conflicts and asked a predefined set of questions. This method usually allows for a detailed insight into subjects' understanding of knowledge and scientific reasoning. Nonetheless, it is often hard to detangle epistemological beliefs from these rich data and this procedure is very time consuming. Thus, most research projects within the *educational* psychology tradition utilize questionnaires that could be administered to groups of subjects and are mostly based on Schommer's (1990) seminal work: Participants are confronted with statements reflecting different epistemological dimensions and have to rate their agreement on Likert-scales. But due to the fact that her proposed dimensions were rarely replicated (see above) most subsequently developed instruments are based on different sets of dimensions. Consider a few examples (partly taken from Stahl & Bromme, in press): the EBI (The Epistemological Beliefs Instrument, Jehng, Johnson, & Anderson, 1993), the EBI (Epistemological Beliefs Inventory, Schraw, Bendixen, & Dunkle, 2002), the DEBQ (Hofer, 2000), or the DBSQ (Buehl, Alexander, & Murphy, 2002) can be considered direct descendants of Schommer's SEQ. Furthermore, more specific instruments that capture beliefs and attitudes about science were proposed, for example, the VASS (Halloun, 1997), the VOSTS (Aikenhead & Ryan, 1992), the SAI-II (Moore & Foy, 1997), as well as specific questionnaires for mathematics (Schoenfeld, 1989), physics (Elby, Frederiksen, Schwarz, & White, 2003; Hammer, 1994), chemistry (Dalgety, Coll, & Jones, 2003) or other science domains (Songer & Linn, 1991).

Alternative formats of assessments are feasible that do not present statements for rating. For example, Alexander and Dochy (1995) developed an instrument that consists of five pre-defined *graphic representations* of the relationship between knowledge and beliefs (e.g., knowledge and beliefs as congruent constructs or beliefs as a component of knowledge). Even though this instrument has a questionnaire format, the scoring of the open answers

for justifying the selections is time-consuming and not feasible for large samples. Another format concerns the use of *adjective-pairs* instead of whole statements. In response to the above sketched weakness of Schommer-type questionnaires, Stahl and Bromme (in press) suggested to conceptually distinguish between denotative and connotative aspects of epistemological beliefs. While all questionnaires of the Schommer-type focus on denotative aspects, the CAEB (connotative aspects of epistemological beliefs, Stahl & Bromme, in press) was developed to capture the more connotative aspects of epistemological beliefs. It has the format of a semantical differential, but with adjectives relating to epistemological issues, and captures two dimensions with 24 antonymous adjective pairs. The dimension *texture* taps beliefs about the structure and accuracy of knowledge (sample item: “structured – unstructured”) and the dimension *variability* refers to beliefs about the stability and dynamics of knowledge (sample item: “dynamic – static”).

Despite the efforts to construct adequate measurement instruments for epistemological beliefs some critical fundamental issues remain – probably unsolvable (Niesse, Vermunt, Abma, Widdershoven, & van der Vleuten, 2004). The first critical issue constitutes a philosophical problem and is relevant to all questionnaires (opposed to open interviews): Forced-choice instruments may be problematic because they constitute a measure of the respondents’ level of agreement with the researchers’ predefined positions (Sandoval, 2005) but not necessarily the best way to capture students’ *own* epistemologies. The second related and probably even more ubiquitous issue concerns the potential inaccessibility of epistemological beliefs. Most people do not regularly discuss epistemological questions such as “What is knowledge?” Thus, individuals’ beliefs systems might be nonexistent, implicit and embryonic or fully-developed and elaborate (Buehl & Alexander, 2001; Schommer-Aikins & Hutter, 2002). Therefore, it is unclear what happens if students without articulated beliefs are forced to fill in an epistemological beliefs questionnaire. A third critical issue concerns the stability of epistemological beliefs: Most theories of epistemological beliefs – at least implicitly – conceptualize epistemological beliefs as relatively consistent and stable “trait-like” theories consistent across learning situations (Hammer & Elby, 2002; Niesse, Vermunt, Abma, Widdershoven, & van der Vleuten, 2004). Based on theoretical as well as based on empirical grounds these unitary assumptions might be faulty. For instance, statements such as “I don’t like movies with an open end” assume that epistemological beliefs towards movies are the same as towards learning situations.

#### *Excursus – Fostering “Sophisticated” Epistemological Beliefs*

The “sophisticated” epistemological view that knowledge is tentative, partially subjective, that it relies upon an empirical basis, that it is socially and culturally embedded, based upon observations and inferences, and that different scientific domains employ different kinds of scientific methods (Sandoval, 2005) can be considered a valuable educational goal (Bromme, 2005; Kardash & Scholes, 1996; De Corte, Op ‘t Eynde, & Verschaffel, 2002;



Schraw, 2001). Sandoval (2005), for example, argued that “lay citizens need to understand science, its power and limits, not because that is good for science [...], but because it is crucial to democracy” (p. 638). Consequently, one line of inquiry within research on epistemological beliefs is concerned with fostering such kind of “sophisticated” beliefs. This strand of research is still in its infancy (Bråten & Strømsø, 2005).

From a theoretical point of view, many instructional methods to foster “sophisticated” epistemological beliefs may seem promising. Qian and Alvermann (2000) suggested four methods: (1) provide students with opportunities to crisscross the landscape of a complex concept and thus examine it from different perspectives (cognitive flexibility theory, Spiro, Feltovich, & Jacobson, 1992), (2) engage students in reflective inquiry (Clarebout & Elen, 2001; Elen & Clarebout, 2000), (3) use images of scientists’ activities from history (Leach, Hind, & Ryder, 2003), or (4) reflect explicitly on teacher’s epistemological objectives (Hammer, 1995). Additionally, Kienhues, Bromme, and Stahl (in press) proposed to introduce controversial issues into the curriculum which may reciprocally foster the development of epistemological beliefs (also Kuhn, 1991; Schommer-Aikins & Hutter, 2002; Mason & Boscolo, 2004; Gill, Ashton, & Algina, 2004). This latter assumption is based on the conceptual change paradigm: A discrepancy between existing beliefs and new experiences may lead to dissatisfaction and change (Hofer, 2004).

Empirical studies investigating the impact of such instructional interventions, however, are rare and so far obtained inconclusive results. Gill, Ashton, & Algina (2004) confronted 161 preservice teachers either with an expository text or with a combination of augmented activation and refutational text. Students who received the refutational text changed their epistemological beliefs in the intended direction; the intervention was successful. Focusing on scientific models from history Leach, Hind, and Ryder (2003) also described promising results: More students demonstrated “sophisticated” beliefs after the intervention. Similar positive effects were obtained by Valanides and Angeli (2005) and Hofer (2004).

Other studies did not obtain such clear-cut positive effects. For example, in a study by Kienhues, Bromme, and Stahl (in press) samples of “naïve” and “sophisticated” students ( $n = 58$ ) were administered either a “neutral” text or a refutational text intended to elicit more “relativistic” beliefs. This short-term intervention affected students’ epistemological beliefs. However, not all changes occurred in the expected direction. Consistently, a mere factual text (“neutral”) resulted in more “naïve” beliefs if effects were obtained at all. Even the “relativistic” instruction was not able to improve students’ pre-instructional “sophisticated” beliefs. Just for the initially “naïve” students the “relativistic” instruction was beneficial, either to prevent more “naïve” epistemologies or to stimulate more “sophisticated” epistemologies. To give another example of inconsistent results, Sandoval and Morrison (2003) attempted to foster more “sophisticated” epistemological beliefs by means of a 4-week inquiry unit on evolution. They interviewed 8 students before and after this interven-

tion. Results do not indicate any significant changes. Similar results were obtained by Clarebout and Elen (2001), Elen and Clarebout (2000) and Tolhurst (2004).

Despite these inconclusive results, instructional attempts to foster more “sophisticated” beliefs should not be neglected. If “sophisticated” beliefs about the nature of scientific knowledge are necessary for normal citizens to make informed decisions about important aspects of their life it is paramount to find a way to convey these beliefs. Additionally, such experimental manipulations facilitate the diagnosis of causality in learning scenarios: When epistemological beliefs are associated with learning processes it is so far unclear if “sophisticated” epistemological beliefs have a positive impact (are the “cause of” positive effects) or if students who employ better learning processes automatically develop more “sophisticated” beliefs (as a “consequence”). One promising way to elicit more “sophisticated” beliefs might be the explicit epistemological discourse (Sandoval & Morris, 2003).

#### *Operationalisations of Epistemological Beliefs*

These selective reviews as well as the excurses above also lead to conclusions with regard to the question how epistemological beliefs can be operationalised empirically. First, as this thesis focuses on the impact of epistemological beliefs on processes of metacognitive calibration, epistemological beliefs need to be measured by economic questionnaires. Based on the general issues and critiques proposed above, it can be concluded that the ideal instrument has not been created yet. Second, adequate instruments were selected based on theoretical considerations (Duell & Schommer-Aikins, 2001; Niessen, Vermunt, Abma, Widdershoven, & van der Vleuten, 2004). Only the core dimensions of epistemological beliefs as defined by Hofer and Pintrich (1997) will be considered: *certainty of knowledge*, *simplicity of knowledge*, *source of knowledge*, and *justification for knowing*. Third, one way to partially protect against potential measurement problems is to employ multiple measures of epistemological beliefs capturing domain-general as well as domain-specific beliefs and capturing denotative as well as connotative aspects: Therefore, domain-general (EBI, Jehng, Johnson, & Anderson, 1993; WKI, Wood & Kardash, 2002) and domain-specific instruments (GCBS, Trautwein & Lüdke, in press) using denotative statements will be supplemented with an instrument using associative connotative adjectives (CAEB, Stahl & Bromme, in press). Fourth, an instructional intervention to elicit more “sophisticated” flexible and evaluativistic epistemological beliefs will be tested to shed some light on the causality of epistemological beliefs: Epistemic doubt (Bendixen, 2002) will be evoked by an *epistemological sensitization* presenting direct comments about the epistemological nature of the content to learners while a control group will not receive such sensitization. Fifth, throughout this thesis the appraising adjectives will be retained: “Naïve” will be consistently used to characterize the detrimental beliefs that knowledge is certain, simple, and can be conveyed by authorities. This is the most “naïve” position that can be measured by questionnaires. And “sophisticated” will be used to characterize the beneficial beliefs that knowledge is uncer-

tain, complex, and should be validated by scientific inquiry (Schommer-Aikins, 2002). This is the most “sophisticated” position that can be measured by questionnaires. Note that most Schommer-inspired questionnaires – including the selected instruments – are not able to capture *evaluativist* views (flexible) epistemologies. However, the epistemological sensitization will try to elicit these kinds of most “sophisticated” epistemologies.

### 2.6.2 Empirical Results and Corresponding Theoretical Explanations

Empirical studies from as different research traditions as argumentation, goal orientation, and learning strategies show a high impact of *epistemological beliefs* on the cognitive and metacognitive components of the learning process as well as on the learning outcome. This is especially true for material of at least moderate complexity.

#### *Results regarding the Preparatory Stages of Self-Regulated Learning*

Indirect evidence for the impact of students’ epistemological beliefs on their *task definitions* comes from research with the reflective judgment model (King & Kitchener, 2002): Students who were confronted with ill-structured problems differed significantly in how they interpreted these problems based on their notions about knowledge and how knowledge could be justified (epistemological beliefs). Students in the *prereflective* stages interpreted these ill-structured problems as well-structured ones: Even though ill-structured problems can be defined as problems where even experts disagree (“Did the Egyptians build the pyramids?”) and which require judgments based on evidence and arguments, these *prereflectional* students thought that these problems had one right answer (“Yes, the Egyptians build the pyramids.”). Students in the *quasi-reflective* and *reflective* level on the other hand were more likely to acknowledge that knowledge pertaining to these ill-structured problems is uncertain (“This question cannot be answered with final certainty. Still, there is a fair amount of evidence against the truth of this statement. Thus, I don’t think that the Egyptians built the pyramids.”). Furthermore, based on the selected research methodology (King & Kitchener, 2002; Kuhn, 1991) it can be inferred that researchers assume an interaction between *task complexity* and the impact of epistemological beliefs: Only ill-structured (and thus more complex) problems were deemed adequate to allow for significant impact of epistemological beliefs. In other words, reflective thinking and epistemological reasoning are assumed to be only “called for when people recognize that some problems cannot be solved with certainty” (King & Kitchener, 2002, p. 38).

With regard to students’ *goal setting* consider the following studies as indirect evidence for the impact of students’ epistemological beliefs: Ryan (1984) classified 90 college students as being either highly *dualistic* or highly *relativistic* based on Perry’s (1970) classification scheme. Furthermore, students were asked to describe how they knew when they understood pre-

sented learning material. Results indicate that the highly *dualistic* students, who believed that knowledge was either right or wrong, reported that they reached sufficient understanding if they could recite the facts. Highly *relativistic* students on the other hand, who believed in relative and context dependent knowledge, reported more criteria for deeper understanding which involved that they reached sufficient understanding if they could apply the learned material to another situation. Therefore, these results indicate that “epistemological beliefs may dictate one’s choice of comprehension standards, and that these epistemological standards, in turn, may control the effectiveness of one’s text processing efforts.” (Ryan, 1984, p. 248). Another study by Bråten and Strømsø (2004) on students’ goal orientation obtained similar results: Students’ concurrently measured mastery goal orientation was predicted by students’ “sophisticated” beliefs in gradual learning (*quick learning*:  $r = -.47$ ,  $p < .001$ ) and dynamic and changing knowledge (*knowledge construction and modification*:  $r = -.45$ ,  $p < .001$ ) while one year later only the belief in gradual learning significantly predicted mastery goals ( $r = -.32$ ,  $p < .01$ ). Furthermore, students performance-avoidance goals were significantly predicted by their “naïve” beliefs in *quick learning* (concurrently:  $r = .31$ ,  $p < .01$ ; after one year:  $r = .44$ ,  $p < .001$ ).

To summarize, these presented results indirectly indicate that more “sophisticated” epistemological beliefs might be related to more accurate *task perceptions* (King & Kitchener, 2002) and with *setting deeper goals* (Ryan, 1984: comprehension standards; Bråten and Strømsø, 2004: mastery goals). Additionally, all self-report questionnaire studies can be interpreted as reflecting students’ *planning*. However, these studies will be reported subsequently as they are most frequently interpreted as capturing students’ *enactment*.

#### *Results regarding the Enactment Stages of Self-Regulated Learning*

First, empirical investigations of students’ *argumentation* will be reported. A study of Kardash and Scholes (1996) is a representative example of a research paradigm introduced by Schommer (1990) that requires students to deal with controversial texts: In this study, ninety-six college students had to read an inconclusive text about the relationship between AIDS and HIV and had to write a concluding paragraph. Furthermore, their epistemological beliefs (SEQ, Schommer, 1990), prior knowledge and their need for cognition were examined. The more students believed in the uncertainty of knowledge (*certainty of knowledge*), the more likely they were to express the inconclusive nature of contradictory evidence ( $r = .32$ ,  $p < .01$ ). On the contrary, students who believed in *certain knowledge* were more likely to misinterpret or ignore contradictory evidence. Furthermore, students’ initial strength of agreement or disagreement with the “HIV causes AIDS” hypothesis was also significantly related to their conclusions ( $r = .25$ ,  $p < .05$ ): Less extreme opinions were positively related to writing conclusions that accurately reflect the inconclusive, tentative nature of the mixed evidence they read. Thus, these results as well as those from other empirical investigations with similar methodology (Mason & Boscolo, 2003; Mason & Scirica, 2006;

Schommer, 1990; Schommer-Aikins & Hutter, 2002) consistently point in one direction: Students with more “sophisticated” beliefs write more adequate conclusions to controversial issues and demonstrate better argumentation.

Second, empirical investigations of students’ *concurrently captured learning strategies* will be reported. For example, Kardash and Howell (2000) investigated concurrent *reading strategies*. Forty students filled in a modified version of the SEQ. Then they read a dual position text about the relationship between HIV and AIDS while thinking aloud. Results indicate that students who believed in gradual learning (*speed of learning*) used more strategies in total ( $r = .40, p < .05$ ), especially more processes and strategies to develop awareness ( $r = .47, p < .01$ ) and to establish intrasentential ties ( $r = .33, p < .05$ ). Students who believed in uncertain knowledge (*certain knowledge*) used more strategies for establishing intersentential ties ( $r = .36, p < .05$ ). Surprisingly, the number of inaccurate text processing was also positively related both to students’ “sophisticated” beliefs that learning is gradual (*speed of learning*:  $r = .41, p < .05$ ) as well as to their “sophisticated” beliefs that knowledge is uncertain (*certain knowledge*:  $r = .32, p < .05$ ). Bromme and Stahl and colleagues on the other hand scrutinized students’ *concurrent help seeking strategies* in a series of studies (Bartholomé, Stahl, & Bromme, 2004; Bartholomé, Stahl, Pieschl, & Bromme, 2006; Bromme & Stahl, 2003). For example, Bartholomé, Stahl, Pieschl, and Bromme (2006) investigated 74 students who worked in 37 dyads during three successive sessions in a regular university course with the hypermedia software “plant identification online”. Students’ epistemological beliefs were captured by the CAEB (Stahl & Bromme, in press). Results indicate an impact of *texture of knowledge*: 9 dyads consistently believing in certain and structured botanical knowledge (“naïve”), 9 dyads believing consistently in uncertain and unstructured knowledge (“sophisticated”) and 19 mixed dyads were compared. Dyads with “sophisticated” beliefs accessed context-sensitive help functions more often than “naïve” dyads. Furthermore, results indicate an impact of *genesis of knowledge*: 12 dyads believing that botanical knowledge is constructed and negotiated (“sophisticated”) were compared with 12 dyads believing that knowledge already exists and just has to be discovered (“naïve”) and 13 mixed dyads. Dyads with consistent “naïve” beliefs made more errors and more follow-up errors (i.e., those dyads were slower to detect their original errors). These results are consistent with the results of the other studies in this series (Bartholomé, Stahl, & Bromme, 2004; Bromme & Stahl, 2003). However, other researchers report a reverse pattern of results with regard to help seeking: Hartley and Bendixen (2003), for example, had 101 students fill in the EBI and subsequently work with a hypermedia tutorial on the bacteria *E. coli*. Results indicate only one effect in the assumed direction: Students “sophisticated” beliefs indicating that they do not accept *omniscient authority* were associated with more non-linear navigation ( $r = -.17, p < .05$ ). All further effects were counterintuitive: Students “naïve” beliefs in *innate ability* were associated with more self-checking ( $r = .27, p < .01$ ) and their “naïve” beliefs in *quick learning* were associated with using the advanced organizer ( $r = .18, p < .05$ ),

the glossary ( $r = .19, p < .05$ ) and the site map more often ( $r = .16, p < .05$ ) as well as with more self-checking ( $r = .16, p < .05$ ). Thus, these results with regard to students' concurrently captured learning strategies indicate in most cases that "sophisticated" epistemological beliefs are related to the usage of more and more adequate learning strategies (Bartholomé, Stahl, & Bromme, 2004; Bartholomé, Stahl, Pieschl, & Brome, 2006; Bromme & Stahl, 2003; Kardash & Howell, 2000; Stadler, 2005); as an exception "sophisticated" beliefs were also associated with less use of help-functions (Hartley & Bendixen, 2003). The results with regard to correctness or accuracy are also inconsistent: While students with "sophisticated" epistemological beliefs used more inaccurate strategies in some studies (Kardash & Howell, 2000); such students were clearly superior in making correct decisions in the studies by Bromme and colleagues.

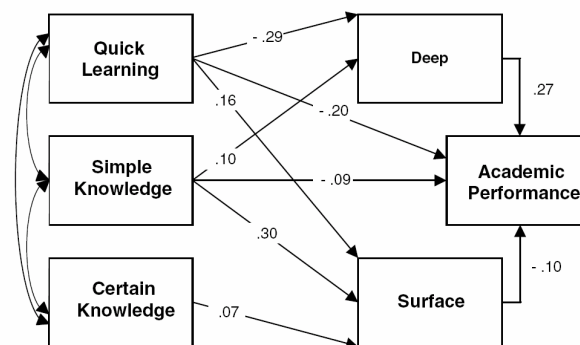


Figure 2.6-1: A path model showing the influence of epistemological beliefs and approaches to learning on academic performance (Figure taken from Cano, 2005, p. 214).

Third, empirical investigations of students' *self-reported learning strategies* (questionnaires) will be reported (Bråten and Strømsø, 2005, 2006; Cano, 2005; Dahl, Bals, & Turi, 2005; Neber & Schommer, 2002; Urhahne & Hopf, 2004; Rozendaal, de Brabander, & Minnaert, 2001). Similar patterns of results were obtained with *interviews* (Tsai, 1998; Whitmire, 2003). For example, in a study by Cano (2005) 1.600 Spanish pupils in three age groups (12 – 14 years, 14 – 16 years, and 16 – 18 years) filled in the Learning Process Questionnaire (LPQ) that measures learning approaches (*deep strategies, surface strategies*) the SEQ (Schommer, 1993: *quick learning, simple knowledge, certain knowledge*) and furthermore their course grades were collected as indicators of their academic performance. Results indicate that students' epistemological beliefs depend on school level and gender: On average, students display more "sophisticated" beliefs on higher school levels, and girls' epistemological beliefs are more "sophisticated" on all school levels than boys'. The impact of epistemological beliefs is depicted in a path model indicating direct and indirect effects (Figure 2.6-1): The belief in *quick learning* was negatively associated with *deep processing*, positively associated with *surface processing* and directly negatively associated with academic performance. The belief in *simple*

*knowledge* was positively associated with *deep processing* as well as with *surface processing*, but directly negatively associated with academic performance. The belief in certain knowledge was only associated with *surface processing*. Furthermore, *deep processing* was positively associated with and *surface processing* was negatively associated with academic performance.

Note, that not all results with regard to students' self-reported learning strategies are consistent with the theoretical assumption that "sophisticated" epistemological beliefs are associated with reporting more deep elaboration strategies and less surface strategies. The effects on the dimension *simple knowledge* are highly contradictory: In some studies "sophisticated" beliefs in complex knowledge are associated with reporting more deep processing strategies (Dahl, Bals, & Turi, 2005), in other studies they are associated with reporting less deep processing strategies (Cano, 2005, see above). In some studies "sophisticated" beliefs in complex knowledge are associated with reporting less surface strategies (Cano, 2005, see above), in other studies they are associated with reporting more surface strategies (Dahl, Bals, & Turi, 2005). For the dimension *certain knowledge* on the other hand, most empirical results confirm the theoretical assumption (Bråten & Strømsø, 2005, 2006; Cano, 2005, Dahl, Bals, & Turi, 2005; Rozendaal, de Brabander, & Minneaert, 2001; Urhahne & Hopf, 2004): More "sophisticated" beliefs in uncertain knowledge are consistently positively associated with strategies of deeper elaboration ("critical processing) while they are negatively associated with surface strategies ("memorizing"). Thus, the results with regard to students' self-reported learning strategies indicate that epistemological beliefs have a significant impact. However, while the effects of the dimension *certain knowledge* are quite consistent with theoretical expectations, the effects of other epistemological beliefs dimensions are less consistent. Furthermore, some studies also indicate no significant impact of epistemological beliefs on students' learning strategies (Neber & Schommer, 2002).

Fourth, consider results with regard to the *learning outcome*. Neither in the studies investigating students' argumentation nor in the studies investigating their concurrent learning strategies consistent positive effects of "sophisticated" epistemological beliefs were found. With regard to *argumentation* tasks, Schommer and colleagues (Schommer, 1990; Schommer, Crouse, & Rhodes, 1992) consistently found positive impact of "sophisticated" beliefs on students' comprehension, while no such effect was detected by Mason and Boscolo (2003). With regard to *concurrent learning strategies*, learners with "sophisticated" beliefs were better at correctly identifying plants (Bartholomé, Stahl, Pieschl, & Brome, 2006) while no impact was detected in other studies (Stadtler, 2005).

To summarize, these presented results indicate significant – but partly inconsistent – impact of epistemological beliefs on written *argumentation* (Kardash & Scholes, 1996; Mason & Boscolo, 2004; Mason & Scirica, 2006; Schommer, 1990), on *concurrent learning processes* such as reading (Kardash & Howell, 2000) and help-seeking strategies (Bartholomé, Stahl, & Bromme, 2004; Bartholomé, Stahl, Pieschl, & Bromme, 2006; Bromme & Stahl, 2003), on *self-reported learning processes* (Bråten & Strømsø, 2005; Cano, 2005; Dahl, Bals, & Turi, 2005;

Neber & Schommer-Aikins, 2002; Rozendaal, de Brabander, & Minnaert, 2001; Schommer-Aikins & Hutter, 2002; Tsai, 1998; Urhahne & Hopf, 2002) and on *learning outcome* (Schommer, 1990, 1993; Schommer, Calvert, Gariglietti, & Bajaj, 1997; Schommer, Crouse, & Rhodes, 1992). Additional non-reported studies indicate significant impact of epistemological beliefs on students' problem-solving with regard to ill-defined problems (Schraw, Dunkle, & Bendixen, 1995), transfer of learning (Jacobson & Spiro, 1995), use of the internet (Bråten & Strømsø, 2006; Bråten, Strømsø, & Samuelstuen, 2006; Hofer, 2004; Stadler, 2005; Whitmire, 2003), interaction with hypermedia (Bendixen & Hartley, 2003; Hartley & Bendixen, 2003; Jacobson, Maouri, Mishra, & Kolar, 1996; Jacobson & Spiro, 1995), learning from constructivist learning environments (Windschitl & Andre, 1998), conceptual change (Mason, 2003; Qian & Alvermann, 1995), understanding of science (Sinatra, Southerland, McConaughy, & Demastes, 2003; Songer & Linn, 1991), preferences for instruction (Tsai, 1998), and engagement in learning (Hofer & Pintrich, 1997).

#### *Results regarding Metacognitive Processes*

First, studies investigating *self-reported metacognitive strategies* will be reviewed. Most results in this strand of research indicate that “sophisticated” beliefs are associated with reporting more metacognitive strategies: Beliefs that success in learning requires cognitive work (*no work*) were related to more “planning” and “monitoring” (Neber & Schommer, 2002), beliefs in complex (*simple knowledge*) and changeable (*certain knowledge*) knowledge were positively associated with more “metacognitive strategies” (Dahl, Bals, & Turi, 2005), beliefs in the validity of experiments for knowledge *justification* and beliefs in changing and *developing* knowledge were positively associated with “metacognitive control strategies” (Urhahne & Hopf, 2004). However, results partially also point in the reverse direction: For example students' “naïve” beliefs in authorities as valid *sources* for knowledge were also positively associated with reporting more “metacognitive control strategies” (Urhahne & Hopf, 2004). These mixed results might be partially explained by the low validity of self-report instruments (Spörer & Brunstein, 2005).

Second, this mixed picture is also found in studies focusing on students' *concurrently captured metacognitive processes*. For example, the above-mentioned study by Kardash and Howell (2000) demonstrated that students' “sophisticated” beliefs in gradual learning (*quick learning*) were positively associated with more strategies for developing awareness during reading. Stadler (2005), however, found no significant relationship between students' epistemological beliefs and their metacognitions during internet research. Bendixen and Hartley (2003) detected no relationship as well. A potential explanation for these mixed effects might be the tasks' simplicity (Bendixen & Hartley, 2003): Some task might have been too simple to detect a significant impact of epistemological beliefs.

Third, two studies from the *traditional calibration paradigm* will be reported: The impact of students' epistemological beliefs on their *calibration* of understanding or their metacompre-



hension was investigated by Schommer (1990) and Schommer, Crouse, and Rhodes (1992). Schommer (1990) had 86 students fill in the SEQ. Subsequently all students were presented with a text passage. As a measure of metacomprehension, students had to rate their confidence in their comprehension on a 4-point scale. Furthermore, they had to write a conclusion for the text passage and had to complete a mastery test. Metacomprehension was diagnosed by the calibration between the students' self-reported comprehension and their performance on the mastery test. Results with regard to students' *calibration* or metacomprehension indicate that the belief in *quick learning* significantly predicted students' overestimation of their understanding of the text passage. Validation of these results comes from a second study: Schommer, Crouse, and Rhodes (1992) had 138 college students fill in the SEQ and subsequently read a text passage under one of two conditions: either to evaluate the comprehensibility of the text or to teach the content to another student. Additionally, students were administered a mastery test to capture their understanding of the texts' content. Results with regard to students' *calibration* or metacomprehension indicate that students' belief in *simple knowledge* was positively associated with their overconfidence. To summarize, students with more "sophisticated" beliefs in gradual learning (*quick learning*; Schommer, 1990) and complex knowledge (*simple knowledge*; Schommer, Crouse, & Rhodes, 1992) were better metacognitively *calibrated*.

To summarize, these presented results mostly indicate significant positive impact of "sophisticated" epistemological beliefs on *self-reported metacognitive processes* (Dahl, Bals, & Turi, 2005; Neber & Schommer, 2002; Urhahne & Hopf, 2003), on *concurrently captured metacognitive processes* (Kardash & Howell, 2000), and on *calibration of comprehension* (Schommer, 1990; Schommer, Crouse, & Rhodes, 1992). However, reverse effects are also detected with regard to self-reported metacognitive processes (Urhahne & Hopf, 2003) and sometimes no effects were found with regard to concurrently captured metacognitive processes (Bendixen & Hartley, 2003; Stadtler, 2005).

### 2.6.3 Conclusion

To investigate the impact of epistemological beliefs on metacognitive calibration processes, this internal condition needs to be validly operationalized to constitute an adequate independent predictor variable. Based on a selective review of theoretical models, potential assessment instruments, and controversial issues a combination of methods was selected: First, different types of economic questionnaires will be used to capture this learner characteristic (EBI, WKI, GCBS, CAEB, see above). Because of the controversially discussed dimensionality of the construct epistemological beliefs, exploratory factor analyses will be computed to obtain meaningful scales (e.g., *certainty of knowledge*). Second, an experimental manipulation of epistemological beliefs will be attempted (*epistemological sensitization*): (1) An

*epistemological introduction* will be administered intended to elicit more “sophisticated” flexible evaluativistic epistemological beliefs, while (2) a *neutral introduction* should cause no change in beliefs. The success of this intervention will be validated by domain-specific questionnaires. Consequently, the impact of epistemological beliefs can be diagnosed by two different methods: If the epistemological sensitization is successful the impact can be diagnosed statistically by comparing these two experimental groups (between-subject). If no epistemological sensitization was implemented or if it failed to elicit the intended effects the scales of the questionnaires can be used as predictor variables.

Based on the reviewed empirical findings the following effects will be expected in the empirical studies of this thesis: First, students with more “sophisticated” epistemological beliefs will excel with regard to most aspects of the self-regulated learning process and with regard to the learning outcome: They will derive more accurate task definitions and goals for deeper elaboration (*preparatory* stage); they will enact more deep elaboration strategies, less surface strategies, and will have better learning outcomes (*enactment* stage); and they will demonstrate more proficient metacognitive skills and calibration (*metacognitive processes*). Second, “sophisticated” epistemological beliefs might not be equally beneficial for learning processes and outcomes in all scenarios; instead an interaction with complexity might be expected: The effects of epistemological beliefs should be more pronounced for more complex material. This is implicitly indicated by material selection (i.e., the use of complex ill-structured tasks and complex materials such as controversial texts). And it is assumed theoretically: “The effects of epistemological beliefs are most obvious in higher order thinking. That is, if students are asked to merely memorize information; there is no need to think mindfully” (Schommer-Aikins & Hutter, 2002, p. 6). However, explicit empirical investigation of this issue so far revealed no significant effects (Schommer, Crouse, & Rhodes, 1992; Mason & Boscolo, 2003). Third, no consistent superiority of students with more “sophisticated” beliefs can be expected because of the inconsistent empirical results. For example, a substantial fraction of results demonstrates a lack of effects or contradictory effects – at least with regard to specific dimensions of epistemological beliefs (see above).

Note that the COPES-model (Winne & Hadwin, 1998; chapter 2.1) predicts similar effects: First, Winne and colleagues explicitly assume that epistemological beliefs impact students’ standards for learning that are used to monitor their whole learning process. Therefore, students with more “sophisticated” epistemological beliefs should excel at most aspects of the self-regulated learning process because they regulate their learning with regard to more adequate goals (Winne, 1996). Second, the descriptions in the COPES-model also imply that more flexible “sophisticated” epistemological beliefs (Elby & Hammer’s, 2001) should also be associated with setting more *accurate* standards. This means that these students’ learning processes should be better calibrated to complexity. To give an example: For a very simple learning task students with “naïve” and “sophisticated” beliefs might not differ much. The “naïve” student might consider all kinds of learning tasks simple because

of her corresponding superficial standards (“Learning is complete if I can recall main concepts.”). A learner with more “sophisticated” epistemological beliefs on the other hand might be better able to diagnose task complexity accurately. Therefore, she should realize that a simple task in fact requires only superficial processing and thus set adequate goals and enact adequate strategies. For more complex tasks, however, these hypothetical students may differ significantly: The “naïve” learner might employ the same superficial learning strategies like memorizing for all kinds of learning tasks. Thus, this learner might only reach a superficial level of understanding and probably might not even realize that this kind of learning is not sufficient for mastering more complex tasks. The “sophisticated” student on the other hand will probably diagnose task complexity accurately and set adequate deep elaboration standards which will also lead to adequate enactment. This student will only be satisfied if deep conceptual understanding is achieved.

Thus, “sophisticated” epistemological beliefs might either have a beneficial (see empirical studies and COPES-model), a detrimental or no *main* effect (see inconsistent results of empirical studies) on the *processes* of self-regulated learning and the *learning outcome*. Instead or in addition to such a main effect, epistemological beliefs might also *interact* with complexity (epistemological beliefs might be more important for more complex tasks or texts).



### 3 Empirical Studies

“Anyway, I don’t think that one could be 100 % certain. But I think that this is an issue of relative certainty. I had the impression that they already understand the reason for what’s going on quite thoroughly, having done many empirical tests, furthermore it’s not something to do with ethics, rather with mechanics and that should be more certain than ethical decisions per se.”

(student DNBH05, study II)

Based on the COPES-model (chapter 2.1), a series of three empirical studies was conducted to investigate learners’ *conditions and processes of metacognitive calibration*. More specifically, learners’ adaptation to two external conditions was scrutinized: *task complexity* (chapter 2.3; studies I and II) and *text complexity* (chapter 2.4; study III). For this endeavor the extended construct definition of “calibration” and the methodology transferred from the traditional calibration paradigm were utilized: *Discrimination* and *relative calibration* were diagnosed and visualized by *calibration graphs* (chapter 2.2). In all studies, the impact of two selected internal conditions was explored: *prior domain knowledge* (chapter 2.5) and *epistemological beliefs* (chapter 2.6). By using this design it was also possible to study potential interactions between these important conditions for self-regulated learning. Furthermore, different studies were devised for the *preparatory* stages of self-regulated learning (study I) and for the *enactment* stages (study II and study III). In the latter studies focusing on the enactment stages it was also possible to investigate if more flexible and accurate calibration in fact was beneficial for the learning outcome as assumed by the COPES-model.

In the first sub-chapter (chapter 3.1) the research questions for all empirical studies will be detailed and in the subsequent sub-chapter (chapter 3.2) the development of the hyper-text on “genetic fingerprinting” will be described. The remaining sub-chapters will each deal with one empirical study: Study I investigates students’ calibration to *task complexity* in the *preparatory* stages of self-regulated learning (chapter 3.3). Study II deals with students’ calibration to *task complexity* in the *enactment* stages of self-regulated learning (chapter 3.4). And study III focuses on students’ calibration to *text complexity* in the *enactment* stages of self-regulated learning (chapter 3.5).

### 3.1 Main Research Questions of the Empirical Studies

#### 3.1.1 Short Review of Research Questions

Most of these research questions have already been discussed in the theoretical chapters, especially in the conclusions. But in order to prepare for the empirical studies of this thesis the most important research questions and arguments will be shortly reviewed.

##### *Do Learners Adapt their Learning to External Conditions?*

The COPES-model (Winne & Hadwin, 1998, chapter 2.1) assumes that one of the core characteristics of self-regulated learning is that learners metacognitively monitor relevant *external conditions* and metacognitively adapt (control) their whole self-regulated learning process accordingly. For example, with regard to task complexity, “students may rehearse information or create mnemonics to meet verbatim reproduction demands, whereas they may generate questions or reorganize learning materials when deep understanding is required” (Broekkamp & van Hout-Wolters, 2006, p. 3; also Pressley et al., 1997). Furthermore, learners who are flexible and accurate in their adaptations to external conditions are assumed to achieve superior learning outcome. Based on a conceptual analysis of hypertext learning two important external conditions were selected: Learners’ adaptations to *task complexity* (chapter 2.3) and to *text complexity* (chapter 2.4) were scrutinized.

However, the COPES-model is underspecified with regard to a detailed description of these adaptation processes. Thus, an adequate terminology and methodology was transferred from the *traditional calibration paradigm* (chapter 2.2). According to this extended construct definition, *calibration* can also denote learners’ adaptation of their whole self-regulated learning process to external conditions. Methodologically, learners’ *discrimination* between different categories of an external condition can be diagnosed by repeated-measure ANOVAs. Learners’ *relative calibration* can be diagnosed by within-subject correlations between learners’ self-regulated learning process and that specific external condition. Additionally, these relations between external conditions and learners’ self-regulated learning processes can be visualized by *calibration graphs*. With one exception it was not possible to determine learners’ *absolute calibration* (absolute accuracy compared with a prescriptive model). Thus, by supplementing the core ideas of the COPES-model (chapter 2.1) with a more detailed terminology and methodology derived from the traditional calibration paradigm (chapter 2.2) the *processes of metacognitive calibration* and their potential benefit for the learning outcome (“*To calibrate or not to calibrate?*”) can be investigated as core research questions of this thesis. More specifically, three research questions can be deduced from these theoretical considerations:

- Do students *discriminate* between tasks and texts of different complexity?
- Do students *calibrate* to tasks complexity and to text complexity?
- Is better calibration to task and text complexity associated with better *learning outcome*?

*Do Internal Conditions Impact these Adaptation Processes?*

Furthermore, the COPES-model (Winne & Hadwin, 1998, chapter 2.1) assumes that *internal conditions* also impact the whole self-regulated learning process and therefore also these adaptation processes. Winne and Hadwin (1998) explicitly discuss *prior domain knowledge* (chapter 2.5) and *epistemological beliefs* (chapter 2.6) as very important internal conditions. Therefore, the impact of these two learner characteristics was scrutinized.

However, the COPES-model is underspecified with regard to a detailed description of the exact way these conditions exert their impact and with regard to potential interactions between internal and external conditions. Internal conditions could have either *beneficial* or *detrimental* effects: For example, prior domain knowledge could facilitate task perception and learners could be better in discriminating between and calibrating to tasks of different complexity. On the other hand, learners with ample prior domain knowledge might fail to notice differences in task complexity because their knowledge automatically compensates for task demands. Thus, they might consider all tasks equally simple demonstrating absent discrimination and calibration. Furthermore, it is unclear if internal conditions have *main* effects independent of external conditions or if they *interact* with external conditions. For example, more prior domain knowledge might be associated with the use of deep elaboration strategies for tasks of all levels of complexity alike. Or more prior domain knowledge might be associated with better calibration of learning strategies to task complexity (i.e., less use of deep processing strategies for simple tasks and more use of deep processing strategies for complex tasks). Thus, not only the processes of metacognitive calibration were explored empirically, but also the impact of relevant conditions (*conditions and processes of metacognitive calibration*). More specifically, the following questions were investigated:

- How exactly do prior domain knowledge and epistemological beliefs impact *discrimination* between tasks and texts of different complexity?
- How exactly do prior domain knowledge and epistemological beliefs impact *calibration* with regard to task complexity or text complexity?
- Do prior domain knowledge and epistemological beliefs impact *learning outcome*?

### *Methodology: Operationalisation of Conditions*

In order to answer these questions, adequate operationalisations for all external as well as internal conditions were derived from corresponding theoretical considerations (chapters 2.3 - 2.6). These operationalisations can be summarized:

- *Task complexity* (chapter 2.3) was operationalised by considering the complexity of the underlying cognitive operations. Thus, the six hierarchical categories of Bloom's revised taxonomy define task complexity (*Bloom-Categories*, in ascending order): (1) *remember*, (2) *understand*, (3) *apply*, (4) *analyze*, (5) *evaluate*, and (6) *create*.
- *Text complexity* (chapter 2.4) was defined by the hierarchical levels of the hypertext on "genetic fingerprinting" (in ascending order): "level 1", "level 2", and "level 3". These texts were constructed by considering semantically differently complex content and this complexity was validated by pilot studies (chapter 3.2.8).
- *Prior domain knowledge* (chapter 2.5) was operationalised by a classical method of the expert paradigm: expert-novice comparison. Humanities students were recruited as novices with regard to "genetic fingerprinting". However, instead of "real" experts advanced students of biology were recruited as discipline experts for this topic (Rouet, Favart, Britt, & Perfetti, 1997). Thus, these quasi-experimental groups can be compared to determine the impact of prior domain knowledge.
- *Epistemological beliefs* (chapter 2.6) were operationalised by two measures: On the one hand they can be measured by questionnaires tapping different aspects of epistemological beliefs (domain-general: EBI, WKI vs. domain-specific: GCBS, CAEB; denotative: EBI, WKI, GCBS vs. connotative: CAEB). On the other hand, an instructional manipulation was devised to elicit more "sophisticated" beliefs (*epistemological sensitization*). If this was successful, the two experimental groups, (1) epistemological introduction and (2) neutral introduction were compared. If this was not implemented or failed, the dimensions of the questionnaires were used as indicators of students' epistemological beliefs.

### **3.1.2 Operationalisation of Research Questions**

As an advanced organizer for the following chapters a short overview of the separate empirical studies will be given including the most important research questions and corresponding operationalisations. In general, very similar research questions were addressed in different studies. Note, that the potentially different impact of conditions in *different stages* can only be determined by comparing study I (*preparatory stage*) and study II (*enactment stage*), both focusing on students' calibration to task complexity. Furthermore, the potentially different impact of *different external conditions* can only be determined by comparing



study II focusing on students' calibration to *task complexity* and study III focusing on students' calibration to *text complexity*.

*Study I: Calibration to Task Complexity in the Preparatory Stages*

Students were presented with six tasks of different complexity (external condition *Bloom-Categories: remember, understand, apply, analyze, evaluate, and create*) and had to give (metacognitive) judgments about important *conditions, operations, standards, and evaluations* for each task. For example, they had to judge the importance of learning strategies for deep elaboration. The following research questions can be answered by this empirical study:

- (1) Do students *discriminate* between *tasks* of different complexity? This question was addressed quantitatively by computing within-subject repeated-measure ANOVAs across all six *Bloom-Categories* for all dependent variables (e.g., the judged importance of deep elaboration learning strategies). Significant effects of the repeated-measure factor *Bloom-Categories* indicate significant discrimination. Additionally, qualitative differences between tasks were explored.
- (2) Do students *calibrate* to *task complexity*? This question was addressed by computing within-subject Gamma correlations between the *Bloom-Categories* and each dependent variable as indices of *relative calibration*. Additionally, students' degree of *absolute calibration* was explored for their own classification of tasks into *Bloom-Categories*.
- (3) The impact of *prior domain knowledge* and *epistemological beliefs* was investigated in two sub-questions for discrimination and calibration:
  - (3a) Are students' prior domain knowledge and their epistemological beliefs associated with students' *discrimination* between tasks of different complexity? This question was addressed by including the quasi-experimental prior domain knowledge groups (humanities students vs. biology students) and the domain-general (EBI) and domain-specific (CAEB) epistemological beliefs scales in the analyses regarding question one.
  - (3b) Are students' prior domain knowledge and their epistemological beliefs associated with students' *calibration* to task complexity? This question was addressed by comparing calibration indices (second research question) between the quasi-experimental prior domain knowledge groups (humanities students vs. biology students) and by correlating these calibration indices with the domain-general (EBI) and domain-specific (CAEB) epistemological beliefs scales.

*Study II: Calibration to Task Complexity in the Enactment Stages*

Students had to actively solve tasks of different complexity (external condition *Bloom-Categories: remember, remember, evaluate, understand, and remember*) with a hypertext on "genetic

fingerprinting”<sup>6</sup>. Simultaneously, their concurrent thoughts and their concurrent actions were captured. Furthermore, their answers to these tasks were collected. The same research questions as for study I can be addressed, but some differences between these studies should be noted: First, while study I only focused on the preparatory stages of learning, study II focused on the enactment stages. Second, not all *Bloom-Categories* were covered in study II. Third, the *epistemological sensitization* was successful in study II, therefore the two experimental groups were compared (epistemological introduction vs. neutral introduction) and causality could be determined. Additionally, a fourth research question with regard to the determinants of learning outcome was added. Thus, the following research questions can be answered by this empirical study:

- (1) Do students *discriminate* between *tasks* of different complexity? This question was addressed quantitatively by computing within-subject repeated-measure ANOVAs across all tasks of different *Bloom-Categories* for all dependent variables (e.g., for students’ concurrent actions). Significant effects of the repeated-measure factor *Bloom-Categories* indicate significant discrimination. Additionally, qualitative differences between tasks were explored (e.g., different navigation strategies).
- (2) Do students *calibrate* to *task complexity*? This question was addressed by computing within-subject Gamma correlations between the *Bloom-Categories* and each dependent variable as indices of *relative calibration*.
- (3) The impact of *prior domain knowledge* and *epistemological beliefs* was investigated in two sub-questions for discrimination and calibration:
  - (3a) Are students’ prior domain knowledge and their epistemological beliefs associated with students’ *discrimination* between tasks of different complexity? This question was addressed by including the quasi-experimental prior domain knowledge groups (humanities students vs. biology students) and the experimental groups of the epistemological sensitization (neutral introduction vs. epistemological introduction) in the analyses regarding question one.
  - (3b) Are students’ prior domain knowledge and their epistemological beliefs associated with students’ *calibration* to task complexity? This question was addressed by comparing calibration indices (second research question) between the quasi-experimental prior domain knowledge groups (humanities students vs. biology students) and between the experimental groups of the epistemological sensitization (neutral introduction vs. epistemological introduction).
- (4) Determinants of the *learning outcome*: Quantitatively, students’ calibration indices indicating students’ adaptation to task complexity were correlated with their perform-

---

<sup>6</sup> For this study learners used a hierarchical hypertext to solve the tasks of different complexity. The hypertext is structured in a way that deeper hierarchical levels present more complex content (study III). However, in this study text complexity is not the primary independent variable, but task complexity is. Text complexity is rather a dependent variable in this design (“Do learners search for deeper level information for more complex tasks?” indicated by logfiles).

ance on the tasks. Additionally, this indicator of learning outcome was compared between groups of different prior domain knowledge (humanities students vs. biology students) and between different experimental groups (epistemological introduction vs. neutral introduction). The impact of further indicators of students' learning strategies was also explored. Additionally, learning processes of the most and least successful students were scrutinized qualitatively.

*Study III: Calibration to Text Complexity in the Enactment Stages*

Students had to “learn as much as possible” about mtDNA analysis with a hypertext chapter about that topic that encompassed three hierarchical levels (external condition *text complexity*: “level 1”, “level 2”, and “level 3”). Simultaneously, their concurrent actions and their judgments regarding text comprehensibility were captured. Furthermore, their learning outcome was determined by a subsequent knowledge test. The same research questions as for study I and study II can be answered, but some differences between studies should be noted: First, students' adaptation to *text complexity* was investigated as main question in this study and not their adaptation to *task complexity*. Second, epistemological beliefs were measured with the domain-general WKI and the domain-specific CAEB.

- (1) Do students *discriminate* between *texts* of different complexity? This question was addressed quantitatively by computing within-subject repeated-measure ANOVAs across all hypertext levels for all dependent variables (e.g., for students' concurrent actions). Significant effects of the repeated-measure factor hypertext levels indicate significant discrimination.
- (2) Do students *calibrate* to *text complexity*? This question was addressed by computing within-subject Gamma correlations between hierarchical hypertext levels and each dependent variable as indices of *relative calibration*.
- (3) The impact of *prior domain knowledge* and *epistemological beliefs* was investigated in two sub-questions for discrimination and calibration:
  - (3a) Are students' prior domain knowledge and their epistemological beliefs associated with students' *discrimination* between texts of different complexity? This question was addressed by including the quasi-experimental prior domain knowledge groups (humanities students vs. biology students) and the domain-general (WKI) and domain-specific (CAEB) epistemological beliefs scales in the analyses regarding question one.
  - (3b) Are students' prior domain knowledge and their epistemological beliefs associated with students' *calibration* to text complexity? This question was addressed by comparing calibration indices (second research question) between the quasi-experimental prior domain knowledge groups (humanities students vs. biology students) and by correlating these calibration indices with the domain-general (WKI) and domain-specific (CAEB) epistemological beliefs scales.

- (4) Determinants of the *learning outcome*: Quantitatively, students' calibration indices indicating students' adaptation to text complexity were correlated with indicators of learning outcome. Additionally, these indicators of learning outcome were compared between groups of different prior domain knowledge (humanities students vs. biology students) and the epistemological beliefs scales (WKI, CAEB) were correlated with these learning outcome indicators. The impact of further indicators of students' learning strategies was also explored. Additionally, learning processes of the most and least successful students were scrutinized qualitatively.

### 3.2 Hypertext on “Genetic Fingerprinting”

The topic of “genetic fingerprinting” within the domain of molecular biology was chosen for a multitude of reasons. First, tasks on all levels of complexity can be constructed easily: On the one hand “genetic fingerprinting” involves well-proven facts such as the structure of DNA. For such facts simple *remember* tasks could be easily constructed. On the other hand, some issues within the topic of “genetic fingerprinting” are discussed controversially and some questions remain open and the corresponding answers uncertain and preliminary. For example, the whole human genome is not sequenced, thus all statements regarding matches and non-matches between two DNA profiles can only give probabilistic answers. For these issues more complex tasks are feasible. Second, these properties of “genetic fingerprinting” also allow for students' different interpretations of tasks and learning materials. For example based, if students possess the “naïve” epistemological belief that knowledge consists of separate and certain facts, they might concentrate on such factual aspects. On the other hand, if students think that knowledge consists of a complex network of interrelated and uncertain bits of knowledge (“sophisticated” view), they might stronger attend to the problems involved in “genetic fingerprinting”. Third, empirical studies support these theoretical claims: Students' epistemological beliefs for the domain of genetics show ample variance (Bromme & Stahl, 2003).

A hypertext on “genetic fingerprinting” was constructed to examine students' learning processes in detail. More specifically, students' adaptation to different levels of *text complexity* could be investigated by observing their processing of hierarchical hypertext levels of different complexity (study III). Differently complex texts were created by describing differently complex content on a semantic level (Kintsch, 1998). Nodes on deeper hierarchical levels contain ascendingly more complex content. These differences also are notable on a surface level: On the top level (“level 1”), all technical terms are explained within the text and simple sentences and primarily motivational and illustrational figures are used. On deeper levels on the other hand technical terms are not directly explained in the text anymore (just in the linked glossary), the nodes are longer, and to adequately represent the

ascendingly more complex content, longer and more nested sentences are used and figures are mostly instructional in nature and also ascendingly more complex.

Furthermore, the hypertext contains different kinds of information to focus on: Students could concentrate on additional “biological background” facts, on additional “examples” that illustrate specific parts of the main hypertext in detail, or they could focus on “problems” that would facilitate the evaluation of the hypertexts’ content. Thus, it can be investigated if students focus on different kinds of information for differently complex tasks (*calibration to task complexity* as investigated in study II). For example, complex tasks might afford critical evaluation, thus it might be appropriate to access hypertext nodes about “problems”. Furthermore, these different kinds of information also allow for an exploration of the impact of students’ epistemological beliefs: It can be investigated if students “sophisticated” epistemological beliefs are in fact associated with focusing on information that explicitly allows for the evaluation of the presented facts like “problem” nodes.

### 3.2.1 MetaLinks – A Template for Hierarchical Hypertexts

The hypertext on genetic fingerprinting was created with MetaLinks (Murray, 2003), an authoring software for hierarchical hypertexts with additional thematic linking. Technically, MetaLinks is FileMaker based and uses Netscape Navigator as a browser. Thus, it uses a server – client architecture. When the hypertext is accessed from a client computer by a Netscape browser, the html code contains a template for the page including JavaScripts. This template is filled with content from relational FileMaker databases. MetaLinks encompasses separate FileMaker databases for “pages” (page title, subtitle, and the main text), “links” (hierarchical links and pre-defined thematic links), “glossary” items (definitions of technical terms), “glossary links” (to specific words on the pages), “media” (figures and tables), “media links” (to specific pages), “user” data (individual logins) and “track” data (all user moves are automatically logged in the so-called logfiles).

The content of these constituent databases is combined to form a typical hypertext node. A typical hypertext node consists of (from top to bottom; Figure 3.2-1): the navigational bar, the page title, the page text including links to glossary definitions of technical terms, figures and tables, links to thematically linked nodes in the pop-up menu “related information”, and links to subordinate nodes at the bottom of the page. MetaLinks allows for some customized layout decisions. In this hypertext on “genetic fingerprinting”, textual information is always given in the left half of the hypertext page, and all figures and tables are presented on the right half.

The image shows a MetaLinks hyperbook interface with three main nodes visible:

- Left Node (mt.2.3.1.1):** Titled "Übersicht über Y-STR-Loci". It contains a table of Y-STR loci with columns for Locus, Repeat, and Allel-Anzahl. Below the table is a detailed text description of Y-STR loci, including a list of 10 loci used in a study and a table of polymorphic STR loci with their repeat numbers.
- Top Node (mt.2.1.1):** Titled "Grundidee der mtDNA-Analyse". It provides a general overview of mitochondrial DNA analysis, including a section on "Biologischer Hintergrund" and "Previously Visited Pages".
- Right Node (mt.3.3.5):** Titled "Beispiel: mtDNA-Analyse bei Spuren an einem Telefonkabel". It describes a forensic case where a telephone cable was used as evidence. It includes a photograph of a cable and a table titled "Ergebnisse der DNA-Analyse" comparing DNA profiles of victims and suspects.

The interface includes navigation elements like "NAV", "History", "Overview", "TOC", "Search", "Parent", "Next/Sibling", and "Logout".

Figure 3.2-1: Sample hypertext nodes: The short bottom hypertext node (mt.2.1.1) is the introductory node to the “basic idea” of the mtDNA analysis (“level 1”), the left node (mt.2.3.1.1) is a “level 3” node from the “basic idea” of Y-STR analysis, and the right node (mt.3.3.5) is an “example” of a forensic case.

### 3.2.2 Navigation in the Hypertext

In order to better understand all navigational options, first an overview of the typical structure of a MetaLinks hyperbook will be given. This structure was also realized in the hyper-text on “genetic fingerprinting”. MetaLinks supports a hierarchical hypertext structure. Such an array of hierarchically organized nodes (Figure 3.2-3) resembles a family tree, thus this analogy is used for navigation: Superordinate nodes are termed “parents”, subordinate nodes are termed “children” and adjacent nodes of the same hierarchical level are called

“siblings” (“previous sibling” for nodes on the left; “next sibling” for nodes on the right). While the labels “parents” and “siblings” appear in the navigational bar at the top of each node, the children of each page are listed at the bottom. Further navigational commands support this hierarchical navigation. “Explain more” at the bottom of a page brings the reader to the first “child” (subordinate node) of that specific page. Clicking “next” is equal to going to the “next sibling”. “Return” is functionally equivalent to the “return” button of a normal browser (it brings the reader to the previously visited hypertext node).

Besides this hierarchical structure and navigational options, MetaLinks promotes multiple ways to navigate to more distant hypertext nodes. For example, MetaLinks allows for further thematic linking between nodes from different hyperbook “chapters” that can be specified by the author. For the hyperbook on “genetic fingerprinting”, three types of thematic links were defined: “biological background”, “examples” and “problems”. A page that possesses such thematic links explicitly tells the reader so. The line “related information exists for this page” is displayed at the bottom of such a page (Figure 3.2-1). These thematic links can be found in the pop-up menu “related information”.

Further navigational options are given in the navigational bar at the top of a node: If a reader knows the name of a specific hypertext node (for example “mt.2.1.1”, Figure 3.2-1), she can type in that number in the “nav” field and click “enter”. Clicking “history” opens an additional window that contains a protocol of the reader’s navigational operations (Figure 3.2-2). By clicking on any of the titles of the previously visited pages, the reader can go back to this specific page (see same is true for all other windows described subsequently). The same function is also given in the “related information” pop-up menu where “previously visited pages” are listed (Figure 3.2-1). The command “glossary” opens the glossary window (an alphabetical list of all glossary entries including the definitions; Figure 3.2-2). “TOC” is the abbreviation for “table of content”. This command opens another window where a hierarchical overview of all nodes is given (Figure 3.2-2). The command “search” also opens a new window (Figure 3.2-2). MetaLinks will search the glossary, the titles and texts for specific terms. Each hit will be reported.

### 3.2.3 The Overall Structure of the Hypertext on “Genetic Fingerprinting”

In the main hierarchical part of the hypertext (Figure 3.2-3), three methods of DNA analysis are detailed: mtDNA analysis, STR analysis, and Y-STR analysis. This overall structure can be either described in terms of content organization or in terms of hierarchy. Additionally, three thematically linked appendices are integrated in the hypertext (Figure 3.2-3) as well as an introduction to genetic fingerprinting (Figure 3.2-3).

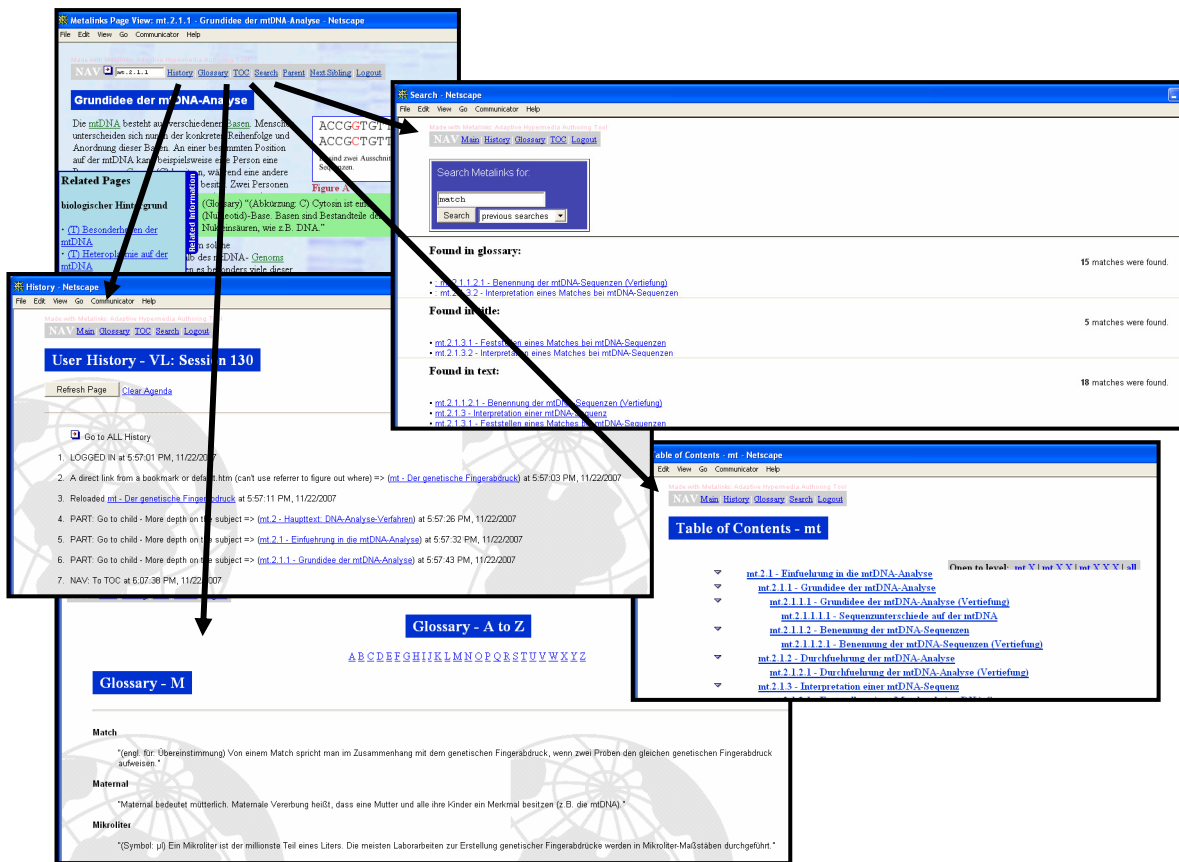


Figure 3.2-2: Windows of a typical MetaLinks hyperbook (“history”, “glossary”, “TOC”, “search”).

With regard to content, similar content is covered in the chapters of the main hierarchical part of the hypertext: First, the “general idea” of the method is described, followed by an introduction to the execution “in the lab”, in introduction to the “interpretation” of results, and an introduction to potential practical “applications” (Figure 3.2-3). More specifically, the first sub-chapter about the “basic idea” generally answers the following questions: What kind of genetic polymorphisms are analyzed (for the STR analysis and the Y-STR analysis descriptions of specific loci are given)? How are the detected inter-individual differences correctly labeled? The second sub-chapter concerns the practical execution of the analysis “in the lab” and answers the question which kind of steps are carried out and which specific laboratory methods are used (e.g., polymerase chain reaction, electrophoresis)? The third sub-chapter deals with the “interpretation” of the obtained results and details how matches or non-matches between two DNA profiles are determined and how a match can be interpreted. In the fourth sub-chapter typical practical “applications” of the methods are given (e.g., forensic cases).



With regard to hierarchy of the main hierarchical part of the hypertext, each of these methods of DNA analysis contains information on three main hierarchical levels: The top-level nodes (“level 1”) are short, simple and serve as introductions (5 nodes for each DNA analysis method: 1 overall introductory node, 4 introductory nodes for the four content areas). By contrast, “level 2” contains nodes of intermediate complexity that elaborate information on the four content areas. The “level 3” contains nodes with very detailed information on certain aspects of these topics.

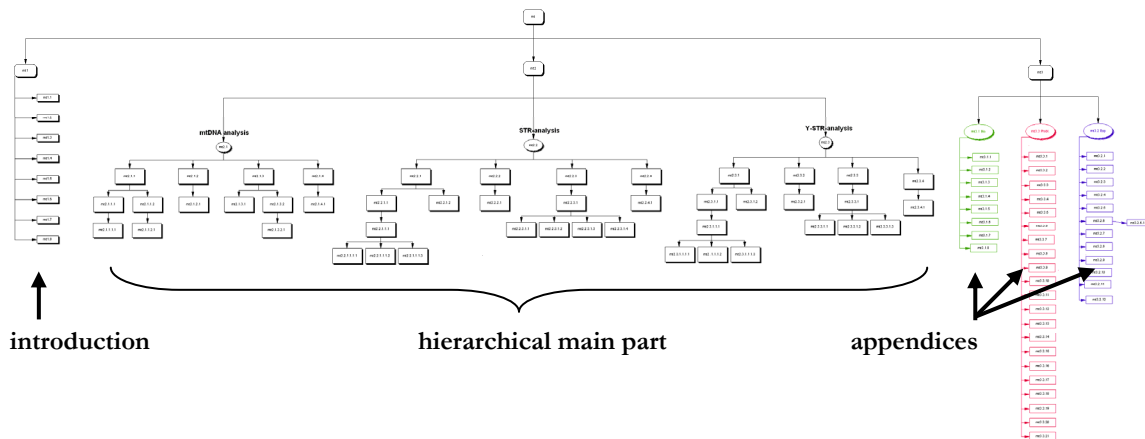


Figure 3.2-3: Overall structure of the hypertext on “genetic fingerprinting”: On the left, the introductory pre-chapter is detailed. In the middle the hierarchical main part of the hypertext is visualized with its three chapters on mtDNA analysis, STR analysis and Y-STR analysis. And on the right side, the three appendices are given (“biological background” (green), “examples” (violet), and “problems” (red)). Only the hierarchical link structure is visualized, all thematic links were not included.

To additionally integrate the nodes on “biological background”, “examples”, and “problems” in this hierarchical structure, three appendices were created (Figure 3.2-3). In the appendix on “biological background” specific topics are introduced that are not usually taught in school. For example, the phenomenon of heteroplasmy on the human mtDNA genome is described. In the appendix on “examples” specific examples, for example practical forensic cases, are given to illustrate the content of the main hierarchical hypertext. In the appendix on “problems” potentials for errors are illuminated (e.g., human errors in the laboratory, inaccurate laboratory methods) as well as inherent uncertainties of the DNA analysis methods (e.g., no match can ever be determined with 100 % certainty). These appendices on “biological background”, “examples”, and “problems” were thematically linked to the main hierarchical hypertext (links not shown in Figure Figure 3.2-3).

To additionally integrate an introduction into the hierarchical structure, a special pre-chapter was created (Figure 3.2-3). This introductory chapter contains similar content as the introductions to genetic fingerprinting that were administered as epistemological sensi-

tization (comparable to the neutral introduction). Basic facts were reported such as the structure of DNA and a description of the human genome.

### 3.2.4 A Quantitative Description of the Hypertext on “Genetic Fingerprinting”

The hypertext on “genetic fingerprinting” contains 106 nodes in total (Figure 3.2-3). Thus, the FileMaker “pages” database also contains 106 entries that encompass the pages’ titles, subtitles and main texts. One hypertext node serves as the first introductory node to the whole hypertext. This node as well as other more organizational hypertext pages does not contain any relevant content. The main *hierarchical* part of the hypertext encompasses 50 nodes (Figure 3.2-3). The top node is just an overview node. The remaining 49 nodes contain relevant content. Fourteen of these nodes detail mtDNA analysis, 18 describe STR analysis and further 17 illustrate Y-STR analysis. The *appendices* contain an almost equal number of nodes. In total 46 nodes form the appendices (Figure 3.2-3). Four of these nodes are organizational in nature and do not contain any relevant content (1 overview of the appendix as a whole and 1 overview node for each of the three appendices). The appendix on “biological background” encompasses 8 contentual nodes, “examples” are illustrated in 13 nodes, and “problems” are scrutinized in 21 nodes. The introduction encompasses 8 nodes with relevant content plus one overview node (Figure 3.2-3).

On average, these nodes are 312 ( $SD = 173$ ) words and 17 ( $SD = 10$ ) sentences long. Furthermore, hypertext nodes contain on average 1.64 ( $SD = 1.62$ ) references and 2.02 ( $SD = 1.32$ ) figures. On average .52 ( $SD = .95$ ) of these figures are primarily motivational in nature and 1.50 ( $SD = 1.21$ ) of these figures are primarily instructional in nature. Additionally, hypertext nodes contain on average 7.75 ( $SD = 4.75$ ) technical terms that are linked to the glossary and .58 ( $SD = .86$ ) thematic links to the appendices.

The hypertexts pages are connected by 193 “links”. One-hundred-and-five of these are *hierarchical* links (Figure 3.2-3). These links are bi-directional. A “parent-child” link (a child that is listed at the bottom of the superordinate page) automatically also functions as a “child-parent” link in the reverse direction (on the subordinate page the command “parent” leads to the superordinate page). The remaining 88 links constitute *thematic* links (not depicted in Figure 3.2-3). These can only be found in the “related information” pop-up menu on the hypertext pages. Thematic links also function bi-directionally. A click on the page title in the pop-up menu “related information” leads to the thematically linked page. A “return” button is automatically created at the bottom of such a page, which leads back to the base page. Fifty-five of these thematic links lead to the “biological background”, 21 to the “problem” nodes and 21 links constitute the thematic links to the “examples”.

Anderson-sequence	(Synonyms: CRS, Cambridge Reference Sequence). The Anderson-sequence constitutes the first completely decoded human mtDNA sequence, first published by Anderson et al. in 1981. After publication, this sequence served as a reference standard, thus it was also called the Cambridge Reference Sequence (CRS). The Anderson-sequence was revised by Andrews et al. in 1991 (so-called revised Cambridge Reference Sequence (rCRS)).
-------------------	---

Figure 3.2-4: Sample glossary entry: In each glossary entry, first synonyms and abbreviations of the technical term are given in brackets. Subsequently the technical term is defined.

The “glossary” database contains the definitions of 166 technical terms. Links to the hypertext pages were created automatically. Each page was searched for the terms contained in the glossary. The first occurrence of a term was linked. Subsequent occurrences of the same terms are not linked to the glossary in order to maintain text readability. In total, the 166 definitions of the glossary are linked to 1378 occurrences of these technical terms in the text (“glossary link” database). Some of these terms are linked to multiple hypertext pages, for example “allele” is linked to 45 pages, “sequence” is linked to 30 pages and “DNA” is linked to 25 pages (this latter number for “DNA” seems quite low, but the actual number is higher because composite words that contain the term “DNA” were linked to the glossary as separate entries, for example “DNA analysis” or “mtDNA”). These technical terms are central to the topic of genetic fingerprinting. Other terms are specific to single hypertext pages and are only linked once, for example “agarose”, “3’-end” or “3’-OH-group”. A specific example of a typical glossary entry is given in Figure 3.2-4.

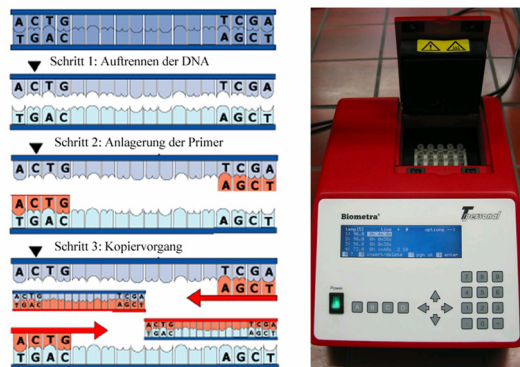


Figure 3.2-5: Sample figures: The right figure illustrates the machinery used for polymerase chain reaction and is *illustratory*, the left figure details the processes on the level of DNA and is *instructional*.

The “media” database of for the hyperbook contains 214 figures and tables. All figures were edited with Photoshop to create figures that are consistent in size and captions. Fur-

thermore, all figures and tables were converted into the .gif format. Some of these figures are primarily *motivational* or illustrational in nature while others are strictly *instructional* (Figure 3.2-5). As all illustrations were designed for a specific hypertext page the “media links” database also contains exactly 214 entries. More specifically, even though some pictures were used multiple times, different captions were inserted to adapt to the context. Thus, these figures are counted as separate entries in the “media” database.

	A	B	C	D	E
1	AJDO27	08.09.2005	12:33:35 PM	LOGGED IN	
2	AJDO27	08.09.2005	12:33:45 PM	LOGIN: To Top Level	mt - Der genetische Fingerabdruck
3	AJDO27	08.09.2005	12:33:52 PM	Go to child	mt.1 - Hintergrundwissen zum genetischen Fingerabdruck
4	AJDO27	08.09.2005	12:37:41 PM	NAV: Parent	mt - Der genetische Fingerabdruck
5	AJDO27	08.09.2005	12:38:15 PM	Go to child	mt.2 - Haupttext: DNA-Analyse-Verfahren
6	AJDO27	08.09.2005	12:38:28 PM	Go to child	mt.2.1 - Einführung in die mtDNA-Analyse
7	AJDO27	08.09.2005	12:40:00 PM	NAV: Parent	mt.2 - Haupttext: DNA-Analyse-Verfahren
8	AJDO27	08.09.2005	12:40:20 PM	NAV: Parent	mt - Der genetische Fingerabdruck
9	AJDO27	08.09.2005	12:40:26 PM	Go to child	mt.1 - Hintergrundwissen zum genetischen Fingerabdruck
10	AJDO27	08.09.2005	12:40:46 PM	Go to child	mt.1.1 - Definition genetischer Fingerabdruck
11	AJDO27	08.09.2005	12:41:40 PM	Return	mt.1 - Hintergrundwissen zum genetischen Fingerabdruck
12	AJDO27	08.09.2005	12:41:45 PM	Go to child	mt.1.2 - Hintergrundwissen: DNA
13	AJDO27	08.09.2005	12:45:19 PM	Return	mt.1 - Hintergrundwissen zum genetischen Fingerabdruck
14	AJDO27	08.09.2005	12:45:24 PM	Go to child	mt.1.3 - Hintergrundwissen: Genom
15	AJDO27	08.09.2005	12:48:02 PM	Return	mt.1 - Hintergrundwissen zum genetischen Fingerabdruck
16	AJDO27	08.09.2005	12:48:07 PM	Go to child	mt.1.4 - DNA-Analyse-Verfahren
17	AJDO27	08.09.2005	12:49:06 PM	Return	mt.1 - Hintergrundwissen zum genetischen Fingerabdruck
18	AJDO27	08.09.2005	12:49:13 PM	Go to child	mt.1.5 - Probengewinnung: Polizei, Recht
19	AJDO27	08.09.2005	12:50:00 PM	Return	mt.1 - Hintergrundwissen zum genetischen Fingerabdruck
20	AJDO27	08.09.2005	12:50:05 PM	Go to child	mt.1.6 - Erstellung eines genetischen Fingerabdrucks im Labor
21	AJDO27	08.09.2005	12:51:06 PM	Return	mt.1 - Hintergrundwissen zum genetischen Fingerabdruck
22	AJDO27	08.09.2005	12:51:11 PM	Go to child	mt.1.7 - Interpretation eines genetischen Fingerabdrucks
23	AJDO27	08.09.2005	12:52:57 PM	Return	mt.1 - Hintergrundwissen zum genetischen Fingerabdruck

Figure 3.2-6: Sample logfile for the user “AJDO27”.

Each user login is captured by the “user” database and each individual user’s navigational history is captured in the “track” database. Thus, the sizes of these databases can not be finally given. These databases grow continuously with the use of the hypertext. Each individual user’s logfile can be easily extracted from the “track” database. Most importantly, it contains the following information: (1) the login of the user, (2) the date and time of the user’s actions, (3) the navigational commands used and (4) the hypertext nodes that were accessed by these commands. Consider the example in Figure 3.2-6. The user “AJDO27” (column A) accessed the hypertext on “genetic fingerprinting” on September the 8<sup>th</sup> 2005 (column B) at 12:33 pm (column C). First, she went to the introduction (mt.1, column E), back to the overall introductory node (mt) and then chose to explore the main hypertext (mt.2), more specifically the sub-chapter on mtDNA analysis (mt.2.1). For this endeavor she used the commands “child” and “parent” (column D).

### 3.2.5 Chapter on mtDNA Analysis

The chapter on mtDNA analysis encompasses 14 nodes that are arranged in a hierarchical structure (Figure 3.2-7). The top two levels encompass five introductory nodes, one overall

introductory node for mtDNA analysis and four introductory nodes for the sub-topics “basic idea” of mtDNA analysis, execution of mtDNA analysis “in the lab”, “interpretation” of mtDNA analysis results and practical “applications” of mtDNA analysis. These nodes are short and serve as introductions (“level 1”). The “level 2” contains six nodes of intermediate difficulty that elaborate information on the four sub-topics. The “level 3” contains three nodes with very detailed information on certain aspects of these sub-topics. Besides this hierarchical structure, the hypertext contains thirty nodes in the appendices that are linked thematically with the main text. They belong to the categories of “biological background” (15 nodes; 2 of these nodes are specific to mtDNA analysis), “examples” (7 nodes) and “problems” (8 nodes). Those nodes differ in length as well as difficulty.

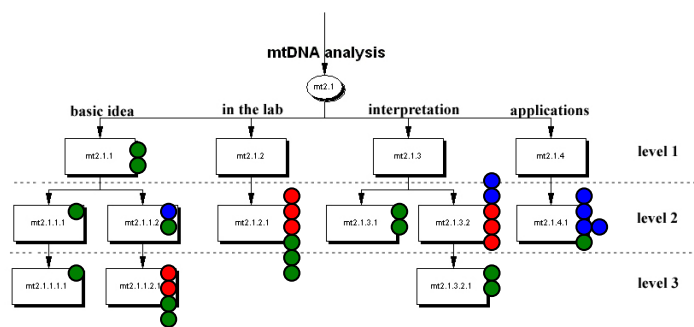


Figure 3.2-7: Visualization of the hypertext chapter on mtDNA analysis. The fourteen nodes are arranged on three hierarchical levels (“level 1 – 3”) and belong to four topics (“basic idea” of mtDNA analysis, execution of mtDNA analysis “in the lab”, “interpretation” of the results of a mtDNA analysis, and practical “applications” of mtDNA analysis). Links to “biological background” nodes are indicated by green circles, “examples” by blue circles and “problems” by red circles.

The content of this chapter on mtDNA analysis was assembled based on textbooks and scientific literature (Anderson, et al., 1981; Andrews et al., 1999; Anslinger & Rolf, 2003; Brinkmann, 2004; Brinkmann & Wiegand, 1997; Brodersen, Anslinger, & Rolf, 2003; Bodowle, Smith, Moretti, & DiZinno, 2000; Carey & Mitnik, 2002; Carracedo et al., 2000; Edson, Ross, Coble, Parsons, & Barritt, 2004; Fourney, 2002; Gill, Sparkes, & Tully, 2001; Handt, Mayer, & Haeseler, 1998; Holland & Parsons, 1999; Isenberg, 2002; Jobling & Gill, 2004; Pfeiffer & Brinkmann, 2000; Rolf & Wiegand, 2004; Salas, Lareu, & Carracedo, 2001; Spencer, 2004; Szibor, Michael, Plate, & Krause, 2000; Tully et al., 2001). In short, mtDNA analysis (analysis of the mitochondrial DNA) involves sequencing of non-coding regions of the human mtDNA genome. Thus, the result of an mtDNA analysis constitutes a sequence of bases (e.g., GAAATGGATGTGTAAAATTTTGAAA, only an exemplary fragment of a sequence is displayed here). As humans possess multiple mitochondria but only one nucleus per cell, mtDNA can often be extracted from degraded cell material, even after this is not possible anymore for nuclear DNA. Thus, this kind of analysis is especially well-suited

for the analysis of ancient DNA. Furthermore, mitochondria are inherited maternally (all individuals of a maternal line possess the same mitochondrial genome; for example a mother and all her children). This property is detrimental for forensic casework, as it does not allow for definite identification of individuals. But the same property is advantageous if applied to genealogy: Genetic relationships between individuals can be tested even if they share only one maternal ancestor a few generations ago.

### 3.2.6 Chapter on STR Analysis

The chapter on STR analysis encompasses 18 nodes that are arranged in a hierarchical structure (Figure 3.2-8). The top two levels encompass five introductory nodes, one overall introductory node for STR analysis and four introductory nodes for the sub-topics “basic idea” of STR analysis, execution of STR analysis “in the lab”, “interpretation” of STR analysis results and practical “applications” of STR analysis. These nodes are short and serve as introductions (“level 1”). The “level 2” contains five nodes of intermediate difficulty that elaborate information on the four sub-topics. The “level 3” contains eight nodes with very detailed information on certain aspects of these sub-topics. Besides this hierarchical structure, the hypertext contains thirteen nodes in the appendices that are linked thematically with the main text. They belong to the categories of “biological background” (3 nodes), “examples” (3 nodes) and “problems” (7 nodes). Those nodes differ in length as well as difficulty.

The content of this chapter on STR analysis was assembled based on textbooks and scientific literature (Benecke, 1997, 2001; Brinkmann, 1998, 2004; Brinkmann & Wiegand, 1997; Brodersen, Anslinger, & Rolf, 2003; Budowle & Allen, 1998; Bodowle, Smith, Moretti, & DiZinno, 2000; Butler, 1998, 2001; Butler, Buel, Crivellente, & McCord, 2004; Butler, Schoske, Vallone, Redman, & Kline, 2003; Butler, Shen, & McCord, 2003; Carey & Mitnik, 2002; Fourney, 2002; Gill, Sparkes, & Tully, 2001; Hohoff & Brinkmann, 2003; Jobling & Gill, 2004; Puers, 2003; Schmitter, 1998; Spencer, 2004; Wiegand & Rolf, 2003). In short, STR analysis involves counting the number of short tandem repeats (STRs, short repetitive sequences like GATA-GATA-GATA) at specific non-coding loci on the human nuclear genome. In order to obtain these results, the DNA at these specific loci is amplified and the length of the DNA fragments is determined. If the length is compared with empirically validated standards, the number of existent STRs can be deduced. This kind of analysis is well suited to identify individuals. Each of the selected STR loci possesses numerous different alleles (and thus is polymorphic). Furthermore, each individual possesses two alleles at each locus, one inherited from the mother, the other inherited from the father. By this mechanism of recombination individuals differ in their alleles at specific STR loci. Thus, if a high number of STR loci are analyzed simultaneously, it is unlikely that two individuals

display the same STR profile by chance. Thus, this method is the most wide-spread method of DNA analysis (e.g., for paternity testing or for forensic cases).

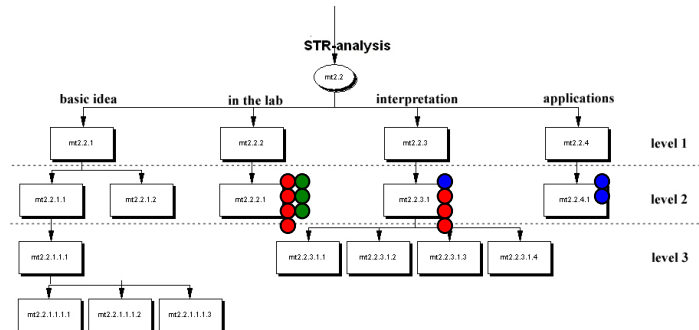


Figure 3.2-8: Visualization of the hypertext chapter on STR analysis. The eighteen nodes are arranged on three hierarchical levels (“level 1 – 3”) and belong to four topics (“basic idea” of STR analysis, execution of STR analysis “in the lab”, “interpretation” of the results of a STR analysis, and practical “applications” of STR analysis). Links to “biological background” nodes are indicated by green circles, “examples” by blue circles and “problems” by red circles.

### 3.2.7 Chapter on Y-STR Analysis

The chapter on Y-STR analysis encompasses 17 nodes that are arranged in a hierarchical structure (Figure 3.2-9). The top two levels encompass five introductory nodes, one overall introductory node for Y-STR analysis and four introductory nodes for the sub-topics “basic idea” of Y-STR analysis, execution of Y-STR analysis “in the lab”, “interpretation” of Y-STR analysis results and practical “applications” of Y-STR analysis. These nodes are short and serve as introductions (“level 1”). The “level 2” contains five nodes of intermediate difficulty that elaborate information on the four sub-topics. The “level 3” contains seven nodes with very detailed information on certain aspects of these sub-topics. Besides this hierarchical structure, the hypertext contains sixteen nodes in the appendices that are linked thematically with the main text. They belong to the categories of “biological background” (8 nodes), “examples” (3 nodes) and “problems” (5 nodes). Those nodes differ in length as well as difficulty.

The content of this chapter on Y-STR analysis was assembled based on textbooks and scientific literature (Benecke, 2001; Brinkmann, 1998, 2004; Brinkmann & Wiegand, 1997; Brodersen, Anslinger, & Rolf, 2003; Bodowle, Smith, Moretti, & DiZinno, 2000; Butler, 2005; Butler, Decker, Kline, & Vallone, 2005; Carey & Mitnik, 2002; Cornacchia & Fitch, 2005; de Knijff, 2003; Gill et al., 2001; Gusmao & Carracedo, 2003; Hohoff & Brinkmann, 2003; Jobling & Gill, 2004; Jobling, Pandya, & Tyler-Smith, 1997; Rolf & Wiegand, 2004; Schoske, Vallone, Kline, Redman, & Butler, 2004; Wiegand & Rolf, 2003). In short, Y-STR

analysis utilizes the same basic idea and methodology as STR analysis. The sole difference is that non-coding regions on the Y-chromosome are analyzed and not on the autosomes. Thus, Y-STR analysis can only be conducted for male individuals. Furthermore, the Y-chromosome is inherited paternally (from the father) as a single haplotype. Thus, all males of the same paternal line possess the same Y-chromosome (e.g., grandfather, father, and son). Therefore, this DNA analysis method does not differentiate well between individuals. Nonetheless, it is useful for special applications, for example for extracting the male portion of evidence in rape cases or for determining ethnic origin.

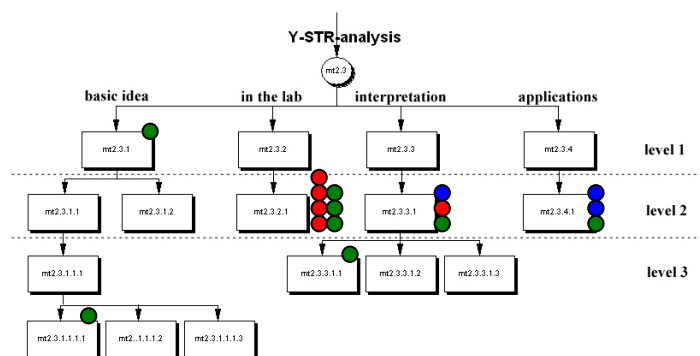


Figure 3.2-9: Visualization of the hypertext chapter on Y-STR analysis. The seventeen nodes are arranged on three hierarchical levels (“level 1 – 3”) and belong to four topics (“basic idea” of Y-STR analysis, execution of Y-STR analysis “in the lab”, “interpretation” of the results of a Y-STR analysis, and practical “applications” of Y-STR analysis). Links to “biological background” nodes are indicated by green circles, “examples” by blue circles and “problems” by red circles.

### 3.2.8 Formative and Summative Evaluation of Text Complexity

The hypertext on “genetic fingerprinting” was extensively evaluated formatively as well as summatively. In total, two purely formative empirical pilot studies and one pilot study with a stronger focus on the summative evaluation of the hypertext were conducted. In (almost) all empirical studies the following research questions were addressed: (1) Can novices like psychology students understand the content of the hypertext? (2) Does *text complexity* vary systematically between different hierarchical levels of the hypertext (i.e., are deeper level nodes more complex)? (3) What has to be done to improve the hypertext?

#### 3.2.8.1 Pilot 1 – Formative Evaluation

During a regular course in psychology, 24 students evaluated first printed prototypes of the texts developed for the chapter on mtDNA analysis. Thirty-one texts were presented, 14 of



these belonged to the three hierarchical levels of the hypertext (5 texts of “level 1”, 6 texts of “level 2” and 3 texts of “level 3”) and further 17 belonged to the appendices on further “biological backgrounds” (2 texts), further “examples” (7 texts) and “problems” (8 texts). Each student in this course evaluated 3 or 4 texts on multiple dimensions: Students noted their reading time and underlined unknown words, thus indicating the number of *lexical* problems (Baker, 1985a; Wagoner, 1983). Furthermore, they underlined the perceived inconsistencies, thus indicating the number of *semantic* problems (Baker, 1985a). Additionally, the students subjectively judged the “simplicity – complexity” and the “structuredness – unstructuredness” of the presented texts (Langer, Schulz von Thun, & Tausch, 1990).

With regard to the first research question (Can novices like psychology students understand the content of the hypertext?) the results are ambiguous. A classroom discussion with all students directly after data collection left the impression that psychology students perceived the presented texts too complex. Nonetheless, students’ judgments (Table 3) reveal that only a few students perceived a few texts very complex. Furthermore, note that these students read the texts without the context given within a regular hypertext (without a glossary and without access to superordinate introductory nodes). When considering these issues, the text seem to be demanding for the psychology students – especially texts planned for deeper hierarchical levels -, but not too complex.

Table 3: Mean values for indicators of *text complexity* in pilot 1

Indices of complexity ( <i>M</i> , <i>SD</i> )	Level 1 ( <i>n</i> = 15)	Level 2 ( <i>n</i> = 18)	Level 3 ( <i>n</i> = 9)
Reading time	1:20 (0:44)	2:51 (1:26)	3:30 (2:57)
Number of lexical problems	1.00 (.85)	1.89 (2.14)	4.44 (4.53)
Number of semantic problems	.60 (1.30)	.72 (.96)	1.11 (1.62)
Judged simplicity <sup>+</sup>	1.67 (.72)	3.06 (1.16)	5.44 (1.33)
Judged structuredness <sup>++</sup>	2.20 (1.42)	2.50 (1.50)	4.11 (1.45)

<sup>+</sup> Students judged simplicity on a 7-point scale from 1 = simple to 7 = complex.

<sup>++</sup> Students judged structuredness on a 7-point scale from 1 = structured to 7 = unstructured.

With regard to the second research question (Does *text complexity* vary systematically between different hierarchical levels of the hypertext?) the following results were obtained: The reading time differed significantly between “level 1” and “level 2” ( $t(31) = -3.66$ ,  $p = .001$ ) and between “level 1” and “level 3” ( $t(22) = -2.72$ ,  $p = .012$ ), but not significantly between “level 2” and “level 3” (Table 3). The number of articulated lexical problems significantly varied between “level 1” and “level 3” ( $t(22) = -2.90$ ,  $p = .008$ ), and varied marginally significantly between “level 2” and “level 3” ( $t(25) = -2.01$ ,  $p = .055$ ), but not between “level 1” and “level 2” (Table 3). The number of mentioned semantic problems did not significantly vary between levels. The explicitly judged simplicity of the texts varied

significantly between “level 1” and “level 2” ( $t(31) = -4.02, p < .001$ ), between the “level 1” and “level 3” ( $t(22) = -3.16, p < .001$ ) and between “level 2” and “level 3” ( $t(25) = -2.65, p = .014$ ). The explicitly judged structuredness of the texts significantly varied between “level 1” and “level 3” ( $t(22) = -3.16, p = .005$ ) and between the “level 2” and “level 3” ( $t(25) = -2.65, p = .014$ ), but not between “level 1” and “level 2”. To summarize, all employed measures of *text complexity* consistently indicate ascending complexity with deeper hierarchical levels, although not all differences were significant. The subjective judgments on the dimension “simplicity – complexity” best differentiated between the three levels. The second research question can be answered affirmative: Texts intended for different hypertext levels did indeed differ significantly in (perceived) complexity.

With regard to the third research question (What has to be done to improve the hypertext?), especially the semantic problems and the perceived structuredness of the texts were considered. Students’ comments were scrutinized and the texts were revised accordingly.

### 3.2.8.2 Pilot 2 – Formative Evaluation

During a second regular course in psychology, 23 students evaluated the revised printed prototypes of the texts developed for the chapter on mtDNA analysis. The 31 presented texts were the same as in the first pilot, albeit in the revised versions. The evaluation procedure was also very similar to the one employed in the first pilot. One change was implemented: Students’ were explicitly told to judge the “simplicity – complexity” and “structuredness – unstructuredness” of the *text* as in the first pilot. Furthermore, a question about the “simplicity – complexity” of the *figures* was added.

With regard to the first research question (Can novices like psychology students understand the content of the hypertext?) the results are similar to those of the first pilot. A classroom discussion with all students directly after data collection left the impression that psychology students perceived the presented texts too complex. Nonetheless, students’ judgments (Table 4) revealed that only a few students perceived a few texts very complex. Furthermore, note that these students read the texts without the context given within a regular hypertext (without a glossary and without access to superordinate introductory nodes). When considering these issues, the text seem to be demanding for the psychology students – especially texts planned for deeper hierarchical levels -, but not too complex.

With regard to the second research question (Does *text complexity* vary systematically between different hierarchical levels of the hypertext?) the following results were obtained: Reading time differed significantly between texts from “level 1” and “level 2” ( $t(23) = -7.152, p < .001$ ) and between texts from “level 1” and “level 3” ( $t(16) = -5.40, p < .001$ ), but not between texts from “level 2” and “level 3”. The number of articulated lexical problems differed significantly between texts from “level 1” and “level 3” ( $t(16) = -5.40, p < .001$ ) and varied marginally significantly between texts of “level 1” and “level 2” ( $t(23)$

= -1.85,  $p = .077$ ), no difference was found between “level 2” and “level 3”. The number of mentioned semantic problems did not show any significant differences. The explicitly judged simplicity differed significantly between texts from “level 1” and “level 3” ( $t(15) = -2.49$ ,  $p = .025$ ). The explicitly judged structuredness did not differ significantly between texts from different levels. Explicitly judged simplicity of figures differed significantly between texts from “level 1” and “level 3” ( $t(12) = -2.31$ ,  $p = .040$ ). To summarize, (almost) all employed measures of *text complexity* consistently indicate ascending complexity with deeper hierarchical levels, although not all differences were significant. The second research question can be answered affirmative: Texts intended for different hypertext levels did indeed differ significantly in (perceived) complexity.

Table 4: Mean values for indicators of *text complexity* in pilot 2

Indices of complexity ( $M$ , $SD$ )	Level 1 ( $n = 11$ )	Level 2 ( $n = 14$ )	Level 3 ( $n = 7$ )
Reading time	1:07 (0:28)	5:04 (1:46)	3:53 (1:37)
Number of lexical problems	.91 (1.14)	2.57 (2.79)	4.71 (3.30)
Number of semantic problems	.91 (1.38)	1.19 (1.66)	1.86 (2.12)
Judged simplicity (text) <sup>+</sup>	3.10 (1.66)	3.85 (1.95)	5.14 (1.68)
Judged structuredness (text) <sup>++</sup>	2.14 (2.19)	2.77 (1.01)	3.00 (1.29)
Judged simplicity (figures) <sup>+</sup>	2.14 (2.19)	3.50 (2.12)	4.71 (1.98)

<sup>+</sup> Students judged simplicity on a 7-point scale from 1 = simple to 7 = complex.

<sup>++</sup> Students judged structuredness on a 7-point scale from 1 = structured to 7 = unstructured.

With regard to research question three (What has to be done to improve the hypertext?), especially the semantic problems and the perceived structuredness of the texts were considered. Students’ comments were scrutinized and the texts were revised accordingly. All underlined words were included in the glossary of this hypertext. The resulting texts were implemented as the chapter on mtDNA in the hypertext on “genetic fingerprinting” and were used for study III (see chapter 3.5).

### 3.2.8.3 Pilot 3 – Summative Evaluation

Twenty-nine students of psychology with a mean age of 23 ( $SD = 22$ ) were recruited to evaluate the comprehensibility and complexity of the hypertext chapters on STR analysis and Y-STR analysis. These were constructed by using the hypertext about mtDNA analysis that was already evaluated extensively as a template (previous chapters). All students evaluated 13 randomly picked hypertext nodes that covered all parts of the hypertext (3 “level 1” nodes, 4 “level 2” nodes, 2 “level 3” nodes, 1 “biological background” node, 1 “example”

node, and 2 “problems” nodes). Students were instructed to try to understand and memorize the main message of each node. During this task, students additionally answered questions about the comprehensibility of the nodes, the text and the figures. Logfiles were collected to capture students’ reading time. After this learning and evaluation phase, students filled in a multiple-choice knowledge test about STR analysis and Y-STR analysis. Half of the students answered the questionnaire by relying on memory alone (free recall), the other half was allowed to re-read the corresponding hypertext node.

With regard to the first research question (Can novices like psychology students understand the content of the hypertext?), the results are satisfactory. Especially the aggregate subjective judgments reveal that students estimate *text complexity* in the mid-complex range (Table 5). Furthermore, comprehensibility was determined objectively: Even these psychology students were able to answer most factual questions about STR analysis and Y-STR analysis correctly, at least if these questions referred to content from “level 1” or “level 2”. For hypertext nodes on “level 3”, the percentage of correctly answered questions indicates still an acceptable level of comprehension. Thus, the hypertext chapters about STR and Y-STR analysis were comprehensible but challenging for psychology students.

Table 5: Mean values for indicators of *text complexity* in pilot 3

Indices of complexity (M, SD)	Level 1 (n = 29)	Level 2 (n = 29)	Level 3 (n = 29)
Reading time	3:43 (1:05)	4:55 (1:20)	4:56 (1:40)
Judgment on text <sup>+</sup>	2.29 (.61)	2.23 (.70)	3.79 (.91)
Judgment on figures <sup>+</sup>	1.84 (.67)	2.34 (.67)	4.06 (1.09)
Total judgment <sup>+</sup>	2.26 (.60)	2.50 (.74)	4.18 (.94)
Percent correctly answered	83.91 (24.59)	59.77 (36.05)	46.55 (42.11)

<sup>+</sup> Students judged comprehensibility on a 7-point scale from 1 = comprehensible to 7 = incomprehensible.

With regard to the second research question (Does *text complexity* vary systematically between different hierarchical levels of the hypertext?), within-subject repeated measure analyses were computed for each indicator of *text complexity* listed in Table 5. All indicators differed significantly between hypertext levels (reading time:  $F(2,27) = 17.76, p < .001$ ; judgments on text:  $F(2,27) = 86.63, p < .001$ ; judgments on figures:  $F(2,27) = 71.35, p < .001$ ; total judgments:  $F(2,27) = 81.94, p < .001$ ; percent correctly answered:  $F(2,27) = 9.91, p = .001$ ). Thus, it can be concluded that students not only subjectively judged hypertext nodes on different hypertext levels differently complex, but furthermore, they needed more time to read more complex nodes and they were also able to answer fewer questions correctly on content from more complex nodes. Consequently, objective and subjective

indicators point in the same direction: Hypertext nodes on deeper hierarchical levels are more complex. Thus, the second research question can be answered affirmative.

### **3.2.8.4 Conclusion**

The results of these formative and summative evaluation studies can be summarized as follows: Across all studies, hypertext nodes on the top two hypertext levels were consistently considered simple to moderately complex. Only if confronted with texts intended for “level 3”, psychology students with no relevant prior domain knowledge considered these texts too complex (e.g., in pilot 1,  $M = 5.44$ ,  $SD = 1.33$ , on a 7-point scale from 1 = simple to 7 = complex). Nonetheless, if “level 3” texts were implemented in the hypertext with all necessary support (superordinate hypertext nodes that introduce the topic and a glossary that explains all technical terms), humanities students judgments indicate decreased perceived complexity (in pilot 3,  $M = 4.18$ ,  $SD = .94$ , on a 7-point scale from 1 = comprehensible to 7 = incomprehensible). Thus, for the formatively evaluated and accordingly revised hypertext research question one can be answered affirmative: Novices like psychology students can understand the content of the hypertext.

With regard to research question two, (almost) all data consistently indicates that deeper hierarchical hypertext levels are in fact more complex. Students reported more lexical and semantic problems with deeper level texts (pilot 1 and pilot 2). These results are verified by more objective data: not only students’ self-reported reading time (pilot 1 and pilot 2), but also the automatically logged hypertext reading time (pilot 3) indicates that students need more time to read deeper-level hypertext nodes. Furthermore, performance data is also consistent: Students perform worse on multiple-choice questions about the content of deeper-level hypertext nodes than about the content of top-level hypertext nodes (pilot 3). Thus, the second research question can also be answered affirmative: *Text complexity* varies between the different hierarchical hypertext levels in the intended direction because deeper level nodes are (perceived) more complex than top-level nodes.

With regard to research question three, the prototypical texts intended for the hypertext (pilot 1 and pilot 2) were revised to eliminate as many of the semantic and structure problems as possible. Furthermore, all reported unknown words (pilot 1 and pilot 2) were included in the hypertexts’ glossary. These revisions seemed to be successful: The summative evaluation data from pilot 3 almost consistently indicates better understanding of the content. The understanding is satisfactory even for humanities students.

### 3.3 Study I - Calibration to Task Complexity in the Preparatory Stages

This study pursued the following research questions: (1) Do students *discriminate* between *tasks* of different complexity (different *Bloom-Categories*)? (2) Do students *calibrate* to *task complexity*? (3) Do the *internal conditions* prior domain knowledge and epistemological beliefs impact these processes of *discrimination* and *calibration*?

Biology students ( $n = 52$ ) and humanities students ( $n = 50$ ) who varied in their epistemological beliefs inspected a set of six tasks from different *Bloom-Categories* and answered the COPES-questionnaire capturing their *preparatory* stages of self-regulated learning for each task. Results indicate that students *discriminate* between tasks of different complexity and *calibrate* their learning process to task complexity (e.g., by judging *superficial processing* learning strategies more important for simpler tasks than for complex tasks). Furthermore, these discrimination and calibration processes were significantly related to their prior domain knowledge and especially to their epistemological beliefs (*internal conditions*). For example, students with “sophisticated” epistemological beliefs judged the use of *deep processing* learning strategies more important across all tasks.

#### 3.3.1 Method

##### 3.3.1.1 Procedure

All students received 10 € reimbursement for participation. To investigate the impact of prior domain knowledge, students were selectively recruited to ensure two levels of knowledge: Biology students ( $n = 52$ ) were recruited during regular courses in biology, humanities students ( $n = 50$ ) were recruited by a posting at the psychological institute.

During the first online session, students filled in questionnaires about their domain-general (EBI) and domain-dependent (CAEB) epistemological beliefs, which took them about 15 minutes (all materials of the online-session can be found in appendices A1 – A4). Sixty-five biology students and 64 humanities students completed the online-questionnaire. The second face-to-face session was held with a maximum of 12 students per session and lasted approximately one hour. Not all students continued: 52 biology students (80% of the original sample) and 50 humanities students (78% of the original sample) participated in this second session. First, students had to answer a short molecular genetics knowledge test. Then, an epistemological sensitization was administered: Students were sorted into two matched sub-samples, one sub-sample received a neutral introduction to genetic fingerprinting, and the other sub-sample received an epistemological introduction in order to

elicit a more “sophisticated” view (the same factual information enriched with comments about the epistemological nature of the presented facts). Subsequently, the domain-dependent CAEB was re-administered as a treatment check of this intervention. In the main part of this session students evaluated six learning tasks that represent the six *Bloom-Categories* of different complexity (*remember, understand, apply, analyze, evaluate, and create*) with the COPES-questionnaire. Tasks were presented in random order. The COPES-questionnaire was used to measure students’ judgments about their *preparatory* stages of self-regulated learning (task definition, goal setting and planning). All material used in this face-to-face session can be found in appendices A5 – A10.

The epistemological sensitization was only marginally successful in this study and thus was excluded from all further analyses. In all analyses regarding the impact of *internal conditions* prior domain knowledge was included as a factor (biology students vs. humanities students) and students’ epistemological beliefs captured by the questionnaires CAEB and EBI were included as covariates.

### 3.3.1.2 Participants

Although the advanced students of biology ( $n = 52$ ) were no experts in the specific topic of genetic fingerprinting they can be considered discipline experts (chapter 2.5.3). These biology students (17 males, 35 females) were on average 22 years old ( $SD = 2.10$ ) and studied in the 3<sup>rd</sup> semester biology or related majors ( $SD = 0.28$ ). They already attended 5 of 10 relevant courses in molecular biology. Adequate background knowledge in molecular biology was verified by the results of a short knowledge test ( $M = 7.23$ ,  $SD = 1.06$ , with 8 points maximum; the knowledge test can be found in appendix A5). Their interest ( $M = 3.85$ ,  $SD = 1.60$ , on a 5-point scale with 1 = very low and 5 = very high) and self-rated prior domain knowledge in molecular biology ( $M = 2.79$ ,  $SD = 0.73$ , on a 5-point scale with 1 = very low and 5 = very high) were also quite high.

Students of humanities ( $n = 50$ ) can be considered novices (chapter 2.5.3). These humanities students (6 males, 43 females) were on average 24 years old ( $SD = 4.47$ ) and studied in the 4<sup>th</sup> semester ( $SD = 2.24$ ) a humanity major such as psychology, history or sociology. They did not attend any of 10 relevant university courses in molecular biology. Low background knowledge in molecular biology was verified by the results of a short knowledge test ( $M = 2.16$ ,  $SD = 1.40$ , with 8 points maximum). Their interest ( $M = 2.94$ ,  $SD = 0.94$ , on a 5-point scale with 1 = very low and 5 = very high) and self-rated prior domain knowledge in molecular biology were also quite low ( $M = 1.90$ ,  $SD = 0.87$ , on a 5-point scale with 1 = very low and 5 = very high).

The difference between biology students and humanities students with regard to their prior domain knowledge was significant on all relevant variables (points in the molecular biology test:  $t(100) = 20.63$ ,  $p < .001$ ; number of attended courses relevant to molecular

biology:  $t(94) = 20.37, p < .001$ ; self-rated prior domain knowledge:  $t(99) = 5.60, p < .001$ ). In all instances biology students displayed more prior domain knowledge. Thus, prior domain knowledge indicated by these quasi-experimental groups (biology students vs. humanities students) will be used as dichotomous independent variable.

### 3.3.1.3 Materials

#### *Epistemological Beliefs Questionnaires*

A combination of two instruments was used to measure epistemological beliefs: The EBI (Epistemological Beliefs Instrument, Jacobson & Jehng, 1999; appendix A3) captures students' domain-general beliefs about the nature of knowledge and knowing by denotative statements. All items that do not refer to epistemology in a strict sense were eliminated (items about learning). The remaining 17 items were distributed as follows: 9 items belong to the original factor *certainty of knowledge* (sample item: "If scientists try hard enough, they can find the answer to almost every question"), 5 items belong to the original factor *omniscient authority* (sample item: "Even advice from experts should be questioned."), and 3 items belong to the original factor *simple view of knowledge* (sample item: "Most words have one clearly defined meaning."). These items were stocked up with 4 items from Wood and Kardash's questionnaire (Wood & Kardash, 2002) and further 2 items that were developed in our lab. Thus, the final version of the EBI administered in this study comprised 23 items. The CAEB (Connotative Aspects of Epistemological Beliefs; Stahl & Bromme, in press; appendixes A2 and A8) had to be completed in reference to the domain of genetics and thus was used to measure domain-dependent epistemological beliefs. This instrument includes 24 pairs of connotative adjectives to measure the dimensions *texture* (structure and accuracy of knowledge; sample item: "structured – unstructured") and *variability* (stability and dynamics of knowledge; sample item: "dynamic – static"). All statements and adjectives in both instruments were rated on 7-point scales.

Because of the revisions of the EBI, an exploratory factor analysis was computed. Although it was not possible to replicate the original factor structure for the EBI, a meaningful solution was obtained: Only one EBI scale was retained (9 items, Cronbach's  $\alpha = .76$ ) that explains 35.91 % of variance and was labeled *EBI-definitude*. This factor measures whether students assume that definite and absolute answers are attainable or whether knowledge is indefinite (sample item: "For most scientific research questions there is only one right answer."). The CAEB was administered twice, once before and once after the epistemological sensitization. The final factor solution for the pre-instructional CAEB comprises the two original factors, *CAEB-texture* (9 items, Cronbach's  $\alpha = .87$ ) and *CAEB-variability* (5 items, Cronbach's  $\alpha = .65$ ) and explains 50.39 % of variance. These factors also elicited satisfactory reliability post-instructionally (*CAEB-texture*: Cronbach's  $\alpha = .82$ ; *CAEB-variability*: Cronbach's  $\alpha = .67$ ). Thus, all subsequent statistical analyses will be con-



ducted with the epistemological beliefs scales *EBI-definitude*, *CAEB-texture* and *CAEB-variability* (the latter ones measured post-instructionally) as predictor variables or covariates.

### *Epistemological Sensitization*

The two introductions served two aims: First, the epistemological introduction was intended to elicit more “sophisticated” epistemological beliefs whereas the neutral introduction should not systematically change students’ beliefs. Second, both introductions should provide all students with a minimum of basic knowledge in molecular genetics (e.g., structure of DNA). Thus, they adequately contextualize students to “genetic fingerprinting”.

Table 6: Excerpts from the epistemological sensitization, differences are highlighted in italics

Neutral introduction	Epistemological introduction
<p>A gene is the basic unit of genetic information. It involves a fragment of the DNA-sequence that contains all information for building proteins. As a gene thus always has a function, DNA-fragments that contain genes are called coding DNA regions. The different conditions of a gene are labeled alleles. For example, a gene for the color of a flower could possess three alleles, red, white and rose.</p> <p><i>To summarize, on this page the concepts genes and alleles were introduced. Furthermore, coding and non-coding DNA were differentiated.</i></p>	<p><i>According to the present state of knowledge, a gene can be considered as the basic unit of genetic information. It involves a fragment of the DNA-sequence that contains all information for building proteins. As a gene thus always has a function – although a mostly unknown one – DNA-fragments that contain genes are called coding DNA regions. The different conditions of a gene are labeled alleles. For example, a gene for the color of a flower could possess three alleles, red, white and rose. Human genes do not usually function in such a simple way. On the contrary, complex interactions between multiple genes – partly also involving non-coding regions – determine human characteristics.</i></p>

In this study, a first prototype of the epistemological sensitization was tested (the German versions of the introductions can be found in appendices A6 and A7). Both versions of the introduction (neutral introduction vs. epistemological introduction) describe the same content which was derived from textbooks and scientific articles (Baron et al., 2004; Brinkmann & Wiegand, 1997; Fridell, 2001; Keller & Hülsmann, 2003; Krawczak & Schmidtke, 1994, 1998; National Research Council, 1996; Spencer, 2004; Weber, 2002) and which was delivered in multiple parts: (1) definition of genetic fingerprinting, (2) biological background about DNA (structure of DNA; coding and non-coding DNA regions; DNA loci and alleles; mutation of DNA), (3) biological background about human genomes (nuclear and mitochondrial), (4) different methods of genetic fingerprinting (RFLP analysis, STR analysis, and mtDNA analysis), (5) police procedures and laws that regulate the attainment of DNA probes, (6) the DNA analysis process in the lab, and (7) the interpretation of results. The layout of these introductions is matched to the design of the hypertext

on “genetic fingerprinting”. Thus, a landscape format is used with the left half of the page dedicated to textual information and the right half dedicated to figures.

Seventy-three comments about epistemological views were inserted in the epistemological introduction in order to sensitize students to such questions (Table 6). These comments differ in length from single words (“*so-called*”) up to multiple sentences. Not all comments stress the *relativistic* view that knowledge is complex, variable and socially constructed, some comments also refer to sound empirical proof for certain facts thus emphasizing an *absolutistic* view. Consequently, the epistemological introduction was intended to foster an *evaluativistic* view of flexible epistemologies (i.e., that some pieces of knowledge have stronger empirical support than others and thus should be recognized as more certain). To counter-balance the increased length of the epistemological introduction, little page summaries were inserted into the neutral introduction. Nonetheless, both introductions still differ in length: The neutral introduction encompasses 18 pages with 2.434 words. The epistemological introduction encompasses 19 pages with 3.030 words.

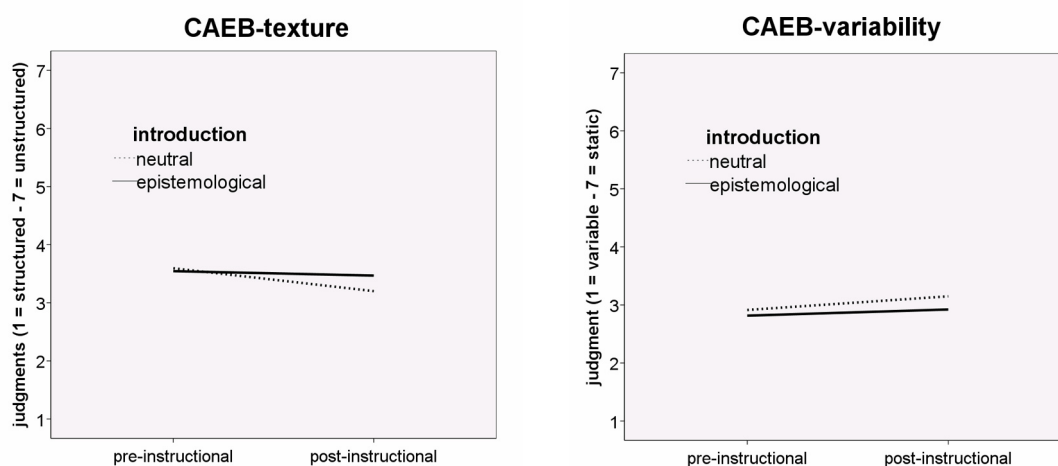


Figure 3.3-1: Students' epistemological beliefs on the scales *CAEB-texture* (left) and *CAEB-variability* (right) as a function of the epistemological sensitization (neutral introduction vs. epistemological introduction).

Empirical results show that the two sub-samples receiving the two versions of the introductions were adequately matched with regard to their epistemological beliefs (no significant pre-instructional differences could be detected: *CAEB-texture*:  $F(1,100) = .07$ ,  $p = .790$ ; *CAEB-variability*:  $F(1,100) = .33$ ,  $p = .565$ ). However, also no consistent effects of the epistemological sensitization could be detected in a repeated-measure analysis: Results indicate a significant multivariate main effect for the repeated-measure factor ( $F(2,99) = 6.50$ ,  $p = .002$ ) that was replicated univariately on both CAEB scales (*CAEB-texture*:  $F(1,100) = 10.59$ ,  $p = .002$ ; *CAEB-variability*:  $F(1,100) = 6.152$ ,  $p = .015$ ). Consistently, students became more “naïve” after reading the introductions (Figure 3.3-1). A differential effect of the two introductions could only be detected for *CAEB-texture* ( $F(1,100)$

= 4.84,  $p = .030$ ): Students with the epistemological introduction became significantly less “naïve” than those with the neutral introduction (Figure 3.3-1, left).

To summarize, this first empirical test of the epistemological sensitization can be considered only marginally successful because on the one hand results indicate trends in the intended direction (post-instructionally students with the epistemological introduction were less “naïve”) but on the other hand these effects were not consistently significant and the epistemological introduction did not have a positive impact (did not elicit “sophisticated” beliefs) but rather seemed to prevent a negative effect: Thus, the subsequent statistical analyses will *not* consider the experimental epistemological sensitization groups. However, the introductions can still be considered beneficial because they served their second purpose well: They contextualized students adequately to the topic of “genetic fingerprinting”.

### *Tasks of Different Complexity*

Bloom’s revised taxonomy (Anderson et al., 2001; chapter 2.3.2) distinguishes between six task classes affording cognitive processes of different complexity (in order of ascending complexity): (1) *remember*, (2) *understand*, (3) *apply*, (4) *analyze*, (5) *evaluate*, and (6) *create*. For this study, one task for each *Bloom-Category* was constructed in a cyclic process.

First, two content experts extensively searched through textbooks about molecular biology and corresponding websites (e.g., Hohoff & Brinkmann, 2004; Keller & Hülsmann, 2003; Krawczak & Schmidtke, 1994, 1998; Spencer, 2004; Wiegand & Rolf, 2003). Furthermore, these experts were introduced to the revised Taxonomy and constructed several tasks representing all *Bloom-Categories*. As a result, a pool of approximately 100 tasks was established, containing molecular biology tasks as well as very specific tasks for the topic of “genetic fingerprinting”. Second, the two content experts as well as three content novices, all deeply familiar with the revised Taxonomy, categorized this pool of learning tasks according to the *Bloom-Categories* in a blind trial. For 39 tasks all five raters immediately agreed, for further 25 tasks only one of the five raters diverged. For the remaining tasks divergence in categorizations was discussed by all five raters and led to rephrasing or deletion of tasks. As result a pool of 86 learning tasks was retained. Third, six tasks for each *Bloom-Category* were chosen by the content experts, resulting in a total of 36 learning tasks. Selection was based on the following criteria: (1) consistently classified tasks were preferred (with high interrater agreement); (2) tasks closely related to genetic fingerprinting were preferred (opposed to tasks related to molecular biology in general); (3) for the *Bloom-Category understand*, the most complex sub-categories like “explain” were ignored because even the authors of the revised Taxonomy (Anderson et al., 2001) consider those sub-categories more complex than the typical *apply* task; (4) overall a balance of content was attempted (not only *evaluate* tasks should deal with problematic aspects of genetic fingerprinting and not only *apply* tasks should deal with algorithms for computing profile matches). These tasks, six for each

*Bloom-Category*, were used in an exploratory study about the *preparatory* stages of self-regulated learning (Stahl, Pieschl, & Bromme, 2006).

Fourth, for this study only one task per *Bloom-Category* was chosen. Selection was based on two requirements: (1) Tasks should be classified as accurately as possible by the students participating in the exploratory study (on the COPEs-item: “classification according to the *Bloom-Categories*”, judgments reaching from 1 = *remember* to 6 = *create*), furthermore (2) tasks should be perceived to be of continuously ascending complexity (on the COPEs-items: “This task is simple.”; “This task is very complex.”; “This task is cognitively demanding.”; all judged on 7-point scales from 1 = totally disagree to 7 = totally agree). Tasks were selected that fulfilled both requirements best. All selected tasks can be found in appendix A10. As an example consider the task of the simplest *Bloom-Category remember*. “DNA is split by which substance? (1) DNA polymerase, (2) *restriction enzymes*, (3) strong acids, (4) ligases, or (5) none of the options is correct (correct option in italics)”.

To conclude, the subsequent statistical analyses will include task complexity as defined by the six tasks representing the six *Bloom-Categories* as independent variable. Note that the same six tasks were administered to all students, but in random order. In all analyses however the order will be depicted in ascending complexity to enhance comprehensibility.

#### *COPEs-questionnaire*

A questionnaire was constructed to examine students’ assessments of the *preparatory* stages of self-regulated learning (*task definition* and *goal setting and planning*). This questionnaire was named COPEs-questionnaire after the model it was derived from (Winne & Hadwin, 1998). The construction of the questionnaire followed a cyclic process as well: First, a pool of 124 potential items was directly deduced from the COPEs-model and from additional articles about metacognition and learning strategies (Anderson et al, 2001; Bloom et al, 1956; Carroll & Korukina, 1999; Broekkamp, van Hout-Wolters, Rijlaarsdam & van den Bergh, 2002; Desoete, Roeyers, & Buysee, 2001; Efklides, Samara & Petropoulou, 1999; Garavalia & Gredler, 2003; Hadwin, Winne, Stockley, Nesbit & Woszycyna, 2001; Krapp, 1993; Luyten, Lowyck & Tuerlinckx, 2001; Nelson, 1999; Nelson & Narens, 1994; Rellinger, Borkowski, Turner & Hale, 1995; Watt, 2004; Wild, 1998; Wild & Schiefele, 1994; Winne & Jamieson-Noel, 2003). Then, five raters who were familiar with this project and the theories of metacognition and learning strategies successively edited this item pool.

The resulting COPEs-questionnaire consists of 46 items (appendix A9). Forty-three of these items have to be rated on 7-point scales, 2 items include short open answers (e.g., judgments of time planned for task completion) and one is a forced-choice classification item. According to the cognitive constituents of the COPEs-model the items of the COPEs-questionnaire are intended to cover students’ opinions about *conditions*, *operations*, *evaluations* and *standards*. More specifically, students had to judge the importance of 10 external (e.g., “sufficient time”) and internal *conditions* (e.g., “motivation”), the importance of

12 broad categories of learning strategies or *operations* (e.g., “memorizing”), and the importance of 18 information (sources) and predictions that indirectly indicate their *standards* (e.g., information sources: “newspapers”, “scientific journals”; types of information: “facts and details”, “contradictory information”; predictions: “estimated time” for task completion). Additionally, students had to rate their agreement to 5 statements (e.g., “This task is easy to solve.”) and fill in a forced-choice item that requires students to classify the presented task into one Bloom-Category to tap their *evaluations*.

Table 7: Results of the exploratory factor analysis of COPES-items across all tasks

Items	factor		
	1	2	3
Sufficient time	<b>,84</b>	,12	-,20
The task is very complex	<b>,82</b>	,15	-,15
Structuring	<b>,79</b>	,13	,06
The task is cognitively demanding	<b>,74</b>	,19	-,10
Integrating	<b>,71</b>	,31	,06
Planning	<b>,70</b>	,20	,11
Elaborating deeply	<b>,58</b>	,47	,04
Ability to draw independent conclusions	<b>,44</b>	,41	-,42
Multiple perspectives	,31	<b>,76</b>	-,21
Newspapers	-,01	<b>,72</b>	-,15
Processing critically	,30	<b>,71</b>	-,31
Contradictory information	,18	<b>,70</b>	,02
Elaborating by discussion	,39	<b>,65</b>	-,00
Memorizing	-,08	-,07	<b>,81</b>
Definitions	,09	-,17	<b>,80</b>

Based on the results of a corresponding exploratory study (Stahl, Pieschl, & Bromme, 2006), 18 items of the COPES-questionnaire were pre-selected for further analyses with regard to students' *discrimination* and *calibration* (i.e., those items that were best suited to capture students' calibration to task complexity). Both open items (“estimated time” and “estimated number of concepts”) as well as the classification item (“classification according to the *Bloom-Categories*”) were included. The remaining 15 items were judged on the same 7-point scales (from 1 = unimportant to 7 = important). These items were subjected to an exploratory factor analysis based on the pooled data from all tasks. The three extracted factors (Table 7) explain 62 % of variance and were labeled *deep processing* (8 items, Cronbach's  $\alpha = .89$ ; sample item: “How important is the strategy of ‘elaborating deeply’ for the solution of the present task?”), *dealing with multiple information sources* (5 items, Cronbach's

$\alpha = .82$ ; sample item: “How important are ‘multiple perspectives’ for the solution of the present task?”), and *superficial processing* (2 items, Cronbach’s  $\alpha = .63$ ; sample item: “How important is the learning strategy of ‘memorizing’ for the solution of the present task?”).

Consequently, all subsequent statistical analyses will utilize these three COPES-factors (*deep processing*, *dealing with multiple information sources*, and *superficial processing*) as well as the three remaining single items (“estimated time”, “estimated number of concepts” and “classification according to the *Bloom-Categories*”) as dependent variables.

### 3.3.2 Results

In this study  $p < .05$  was defined as significant and it was further differentiated between highly significant with  $p < .01$  and  $p < .001$ .

#### 3.3.2.1 Do Students Discriminate Between Tasks of Different Complexity?

To answer the first research question, within-subject repeated-measure analyses across tasks from all six *Bloom-Categories* were calculated for each dependent variable. To enrich these quantitative analyses with more qualitative data, students’ judgments for tasks representing single *Bloom-Categories* were scrutinized as well as the corresponding *calibration graphs*.

##### *Quantitative Analyses*

A within-subject repeated-measure MANOVA was calculated for the three COPES-factors. Results indicate a highly significant multivariate main effect for the repeated-measure factor *Bloom-Categories* ( $F(15,87) = 60.22, p < .001$ ) which was replicated on each COPES-factor separately (*deep processing*:  $F(5,505) = 148.38, p < .001$ ; *dealing with multiple information sources*:  $F(5,505) = 169.86, p < .001$ ; *superficial processing*:  $F(5,505) = 62.78, p < .001$ ). Furthermore, within-subject repeated-measure ANOVAs were computed for each of the remaining single items. Results also indicate significant effects of the repeated-measure factor *Bloom-Categories* (“estimated time”:  $F(5,97) = 17.15, p < .001$ ; “estimated number of concepts”:  $F(5,94) = 19.66, p < .001$ ; “classification according to the *Bloom-Categories*”:  $F(5,90) = 258.06, p < .001$ ).

##### *Qualitative Analyses*

Only exemplary results will be detailed with regard to the three COPES-factors. Results indicate that only extremely simple or extremely complex tasks (*remember* or *create*) elicited extreme judgments (differing significantly from the scale mean; Figure 3.3-2): For the simplest *remember* task, students judged *deep processing* ( $M = 2.61, SD = 1.18$ ) and *dealing with*

*multiple information sources* ( $M = 2.28$ ,  $SD = 1.03$ ) extremely unimportant, but *superficial processing* ( $M = 5.52$ ,  $SD = 1.51$ ) extremely important. For the most complex *create* tasks almost the reverse picture emerged. Students judged *dealing with multiple information sources* extremely important ( $M = 5.56$ ,  $SD = .92$ ), while they judged *superficial processing* extremely unimportant for the *create* task ( $M = 2.71$ ,  $SD = 1.18$ ). For the moderately complex tasks on the other hand (*understand* through *evaluate*) only moderate mean judgments were given. More specifically, the COPES-factor *superficial processing* elicited almost equal judgments for all moderately complex tasks (ranging from  $M = 3.95$  ( $SD = 1.70$ ) for *understand* to  $M = 4.21$  ( $SD = 1.28$ ) for *analyze*). *Deep processing* on the other hand was judged to be of quickly ascending importance for more complex tasks, but judgments reached a plateau starting with *analyze* ( $M = 5.23$ ,  $SD = .93$ ). The judgments for the COPES-factor *dealing with multiple information sources* were significantly lower for the moderately complex tasks, but of gradually ascending importance for the tasks *apply* through *create*.

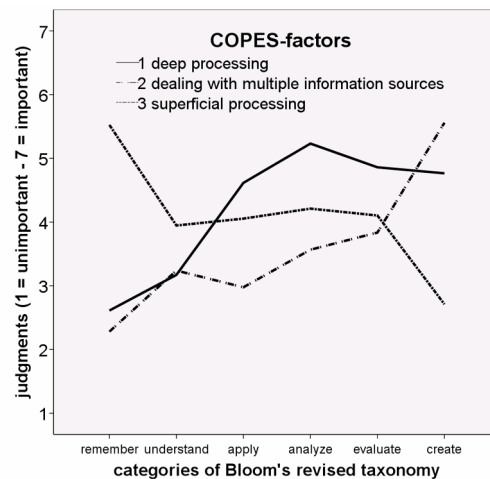


Figure 3.3-2: Calibration graphs depicting students' judgments on the three COPES-factors as a function of task complexity (*Bloom-Categories*).

To summarize the results with regard to research question one: Students' aggregate judgments captured by the COPES-factors, their answers to open items and to the classification item demonstrate significant *discrimination* between tasks of different complexity. This is supported by quantitative as well as more qualitative analyses.

### 3.3.2.2 Do Students Calibrate to Task Complexity?

To answer the second research question measures of relative and – in one case – absolute calibration were utilized: For all dependent variables Goodman-Kruskal Gamma correlations ( $G$ ) between students' judgments on each dependent variable and the *Bloom-Categories*

were computed ( $n = 6$ , for six categories) and subsequently Z-transformed into indices of *relative calibration*. *Absolute calibration* was determined for the item “classification according to the *Bloom-Categories*”. All calibration indices were tested against zero and scrutinized with regard to their effect sizes.

Indices of *relative calibration* for all dependent variables significantly differ from zero and mostly correspond to correlations of at least large effect size ( $G > .50$ ). More specifically, the mean calibration index for the COPES-factor *deep processing* is .60 ( $SD = .41$ ) which corresponds to a correlation of  $G = .54$  and significantly differs from zero ( $t(101) = 14.69$ ,  $p < .001$ ). This positive association between the *Bloom-Categories* and *deep processing* indicates that students judge *deep processing* to be quite unimportant for simple tasks and of ascending importance for more complex tasks (Figure 3.3-2). For the COPES-factor *dealing with multiple information sources*, the picture is similar (mean calibration index = .95,  $SD = .70$ ; corresponds to  $G = .74$ ; significantly differs from zero:  $t(101) = 13.70$ ,  $p < .001$ ; Figure 3.3-2). For the COPES-factor *superficial processing* the picture is reversed: The mean calibration index is -.61 ( $SD = .88$ ) which corresponds to a correlation of  $G = -.55$  and significantly differs from zero ( $t(101) = -7.03$ ,  $p < .001$ ). This negative association between the *Bloom-Categories* and *superficial processing* indicates that students on average judge *superficial processing* to be quite important for simple tasks and of descending importance for more complex tasks (Figure 3.3-2). For the three remaining single items, significant positive calibration indices were found (“estimated time”: mean calibration index = .65,  $SD = .55$ ; corresponds to  $G = .57$ ; significantly differs from zero:  $t(101) = 11.96$ ,  $p < .001$ ; “estimated number of concepts”: mean calibration index = .51,  $SD = .72$ ; corresponds to  $G = .47$ ; significantly differs from zero:  $t(101) = 7.13$ ,  $p < .001$ ; “classification according to the *Bloom-Categories*”: mean calibration index = 1.23,  $SD = 1.02$ ; corresponds to  $G = .85$ ; significantly differs from zero:  $t(101) = 12.23$ ,  $p < .001$ ).

Furthermore, *absolute calibration* was diagnosed for students’ “classifications according to the *Bloom-Categories*”: Students’ judgments could be directly compared to the correct solution (the correct classification of tasks). Students on average classified 47.17 % of the six tasks correctly ( $M = 2.83$ ,  $SD = 1.26$ ). The corresponding *calibration graph* (Figure 3.3-3) reveals that students slightly overestimate the complexity of simpler tasks (*remember – apply*) while they underestimate the complexity of more complex tasks (*analyze – create*) if their judgments are compared with a hypothetical “line of perfect calibration”.

To summarize the results with regard to research question two: Students aggregate judgments captured by the COPES-factors, their answers to open items and to the classification item demonstrate significant *calibration* with regard to *task complexity*, both if measures of *relative calibration* are considered and if measures of *absolute calibration* are considered.



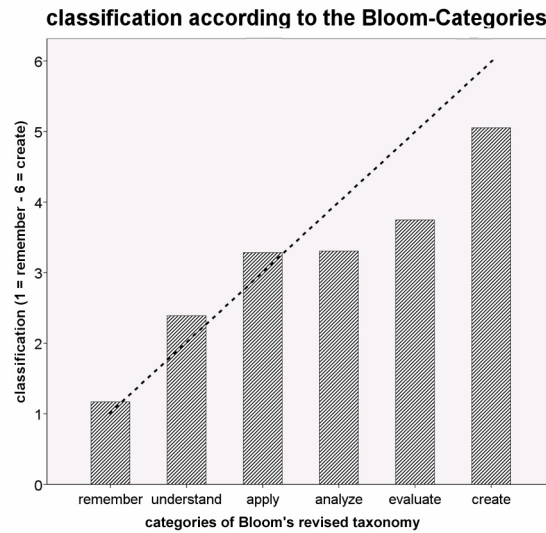


Figure 3.3-3: Calibration graph depicting students' "classification according to the *Bloom-Categories*" as a function of task complexity (*Bloom-Categories*). The dotted line represents the "line of perfect *calibration*".

### 3.3.2.3 The Impact of Prior Domain Knowledge and Epistemological Beliefs

Before reporting the relation between the *internal conditions* – in this case, students' prior domain knowledge (biology students vs. humanities students) and students' epistemological beliefs (indicated by their values on the scales *EBI-definitude*, *CAEB-texture* and *CAEB-variability*) – and students' *discrimination* and *calibration*, some descriptive values for the epistemological beliefs factors will be covered and all relations between these predictor variables will be detailed.

Students in general tended to believe more in indefinite and vague knowledge, thus displaying a "sophisticated" view on the domain-general EBI scale (*EBI-definitude*:  $M = 2.85$ ,  $SD = .81$ ; on a 7-point scale from 1 = indefinite to 7 = definite). Additionally, students tended to believe a little more in structured knowledge in molecular genetics pre-instructionally, thus displaying a mildly "naïve" epistemological view with regard to *CAEB-texture* ( $M = 3.57$ ,  $SD = .95$ ; on a 7-point scale from 1 = structured to 7 = unstructured). Post-instructionally, students displayed a more "naïve" view on *CAEB-texture* ( $M = 3.22$ ,  $SD = .80$ ). Furthermore, students tended to believe in dynamic and flexible knowledge in molecular genetics pre-instructionally, thus displaying a "sophisticated" epistemological view on *CAEB-variability* ( $M = 2.86$ ,  $SD = .85$ ; on a 7-point scale from 1 = flexible to 7 = inflexible). Post-instructionally, students tended to believe less in dynamic and flexible knowledge in molecular genetics (*CAEB-variability*:  $M = 3.04$ ,  $SD = .85$ ).

Correlational analysis revealed that *CAEB-texture* was significantly related to *CAEB-variability* ( $r = -.52$ ,  $p < .001$ ): Students who believed in unstructured knowledge in genetics also believed in relative knowledge in genetics. To also account for prior domain knowl-

edge differences, epistemological beliefs on all scales were compared between biology and humanities students. Only one effect was detected: Humanities students believed more in indefinite knowledge than biology students (*EBI-definitude*:  $t(100) = 3.25, p = .002$ ).

*Are students' prior domain knowledge and their epistemological beliefs associated with their discrimination between tasks of different complexity?*

To answer this research question, students' prior domain knowledge was included as a dichotomous factor (biology students vs. humanities students) and students' epistemological beliefs were included as covariates (*EBI-definitude*, *CAEB-texture*, and *CAEB-variability*) in the within-subject repeated-measure analyses across the six *Bloom-Categories*. To visualize significant results, *calibration graphs* were created. For the epistemological beliefs scales, this visualization was achieved by median-split.

With regard to the effects of *prior domain knowledge*, the MANCOVA for the three COPES-factors indicates a significant multivariate interaction between the *Bloom-Categories* and prior domain knowledge ( $F(15,83) = 2.03, p = .022$ ) that was univariately only replicated significantly on the COPES-factor *deep processing* ( $F(5,485) = 2.94, p = .013$ ; Figure 3.3-4, left): Biology students judged *deep processing* to be of ascending importance from *remember* tasks through *analyze* tasks and their judgments reached a plateau for *analyze*, *evaluate* and *create* tasks. Humanities students did not discriminate on such a fine-grained level. They judged *deep processing* to be quite unimportant for *remember* and *understand* tasks and quite important for all more complex tasks.

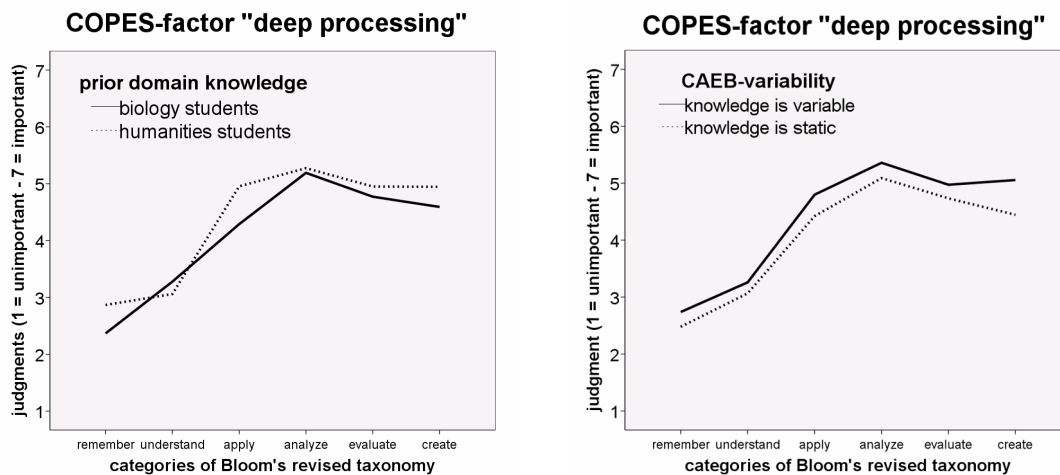


Figure 3.3-4: Calibration graphs depicting students' judgments on the COPES-factor *deep processing* as a function of task complexity (*Bloom-Categories*) and prior domain knowledge (biology students vs. humanities students, left) or *CAEB-variability* (median-split, right).

With regard to the effects of *epistemological beliefs*, the MANCOVA for the three COPES-factors revealed the following effects: A significant multivariate main effect of *CAEB-variability* ( $F(3,95) = 2.85, p = .042$ ) was univariately replicated significantly on the COPES-factors *deep processing* ( $F(1,97) = 5.23, p = .024$ , Figure 3.3-4, right) and *dealing with multiple information sources* ( $F(1,97) = 7.576, p = .007$ , Figure 3.3-5, left): Students, who considered knowledge in genetics variable (“sophisticated” view on *CAEB-variability*) also considered *deep processing* and *dealing with multiple information sources* to be more important across all tasks than more “naïve” students – regardless of task complexity. Furthermore, a significant univariate interaction between *CAEB-variability* and the repeated-measure factor *Bloom-Categories* was detected on the COPES-factor *dealing with multiple information sources* ( $F(5,485) = 2.34, p = .041$ , Figure 3.3-5, left), indicating that the above-mentioned main effect was most significant for the *Bloom-Categories* *remember* through *analyze*, while it disappeared for the more complex tasks *evaluate* and *create*. Additionally, the ANCOVA for the item “classification according to the *Bloom-Categories*” also indicated a significant main effect of *CAEB-variability* ( $F(1,90) = 4.59, p = .035$ , Figure 3.3-5, right): More “sophisticated” students who believed in variable knowledge in genetics classified tasks in more complex *Bloom-Categories* (especially *analyze* tasks).

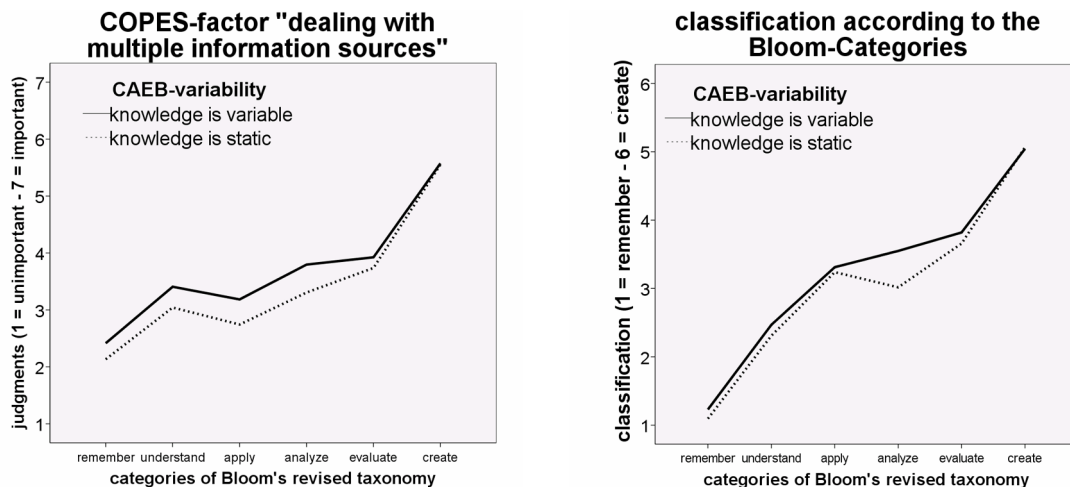


Figure 3.3-5: Calibration graphs depicting students' judgments on the COPES-factor *dealing with multiple information sources* (left) and on students' “classifications according to the *Bloom-Categories*” (right) as a function of task complexity (*Bloom-Categories*) and *CAEB-variability* (median-split).

To summarize the results with regard to research question three and *discrimination: Prior domain knowledge* elicited only one effect (indicating that biology students discriminated on a more fine-grained level) while *epistemological beliefs*, especially those measured by *CAEB-variability*, elicited multiple effects: “Sophisticated” beliefs on this scale were consistently associated with judging indices of deep elaboration more important across all tasks.

*Are students' epistemological beliefs and their prior domain knowledge associated with their calibration to task complexity?*

To answer this research question, students' epistemological beliefs on all scales (*EBI-definitude*, *CAEB-texture*, and *CAEB-variability*) were correlated with their *calibration* indices for all dependent variables (COPES-factors and remaining single items). To also account for the impact of prior domain knowledge, the *calibration* indices of biology students and humanities students were compared via *t*-tests.

With regard to *prior domain knowledge*, only one significant difference was found: Biology students displayed significantly better calibration with regard to "estimated number of concepts" than humanities students ( $t(100) = 2.09, p = .039$ ; biology students: mean calibration index = .65 ( $SD = .83$ ); humanities students: mean calibration index = .36 ( $SD = .56$ )). With regard to *epistemological beliefs*, only one significant correlation was found: "Naïve" beliefs in the definitude of knowledge in general (*EBI-definitude*) were associated with better calibration with regard to "estimated number of concepts" ( $r = .26, p = .009$ ).

To summarize these results with regard to research question three and *calibration: Prior domain knowledge* was beneficial for calibration while "sophisticated" *epistemological beliefs* were detrimental for calibration.

### 3.3.3 Local Discussion

The main goal of this study was to replicate and further explore the empirical findings of the corresponding exploratory study (Stahl, Pieschl, & Bromme, 2006). Students evaluated six learning tasks of different *Bloom-Categories* with the COPES-questionnaire. Thus, this study thoroughly investigated the first two *preparatory* stages of the COPES-model.

#### 3.3.3.1 Discrimination and Calibration

Students consistently and significantly demonstrated their ability to adapt their judgments to the complexity of the learning tasks of different *Bloom-Categories*: The analyses concerning students' *discrimination* and *calibration* consistently indicate positive results.

More specifically, analyses with regard to students' *discrimination* indicate significant differences in students' judgments between tasks of different *Bloom-Categories* for all dependent variables (COPES-factors and remaining single items). With regard to students' *calibration* to task complexity (*Bloom-Categories*), students demonstrated good *relative calibration*: All calibration indices differ significantly from zero and mostly correspond to correlations of at least large effect size. All analyses consistently indicate that students judge indicators of deep processing more important for more complex tasks while they judge indicators of superfi-

cial processing less important for more complex tasks. Additionally, students' good *absolute calibration* was demonstrated for the item "classification according to the Bloom-Categories". Thus, students metacognitively monitored task complexity and gave their judgments accordingly.

Still, one issue should be noted: More fine-grained analyses revealed that students were significantly better at discriminating between distant *Bloom-Categories* than they were at discriminating between adjacent *Bloom-Categories* (e.g., compare *superficial processing* between *understand* and *apply* tasks, Figure 3.3-6). This pattern is hardly surprising, given that even the authors of Bloom's revised taxonomy (Anderson et al., 2001) admit that categories may overlap. Additionally, Bloom's revised taxonomy does not propose a prescriptive model of what kind of *task definitions, goals and plans* are adequate for which type of task. Probably, students' judgments represent an adequate fit for the presented tasks.

These results are consistent with the COPES-model (Winne & Hadwin, 1998; chapter 2.1) that assumes that students systematically adapt their learning process to external conditions. Additionally, these results are partly consistent with those of previous empirical studies about task complexity (chapter 2.3: Gall, 2006; Klayman, 1985; Rouet, 2003; Winne & Jamieson-Noel, 2003). These studies mostly focused on the *enactment* of learning strategies and consistently indicate that learners demonstrate good self-regulation for simple tasks but less adequate self-regulation for complex tasks. The results of this study are consistent because in both cases learners process differently complex tasks differently and systematically adapt their behavior or their judgments to task complexity. However, the results of this study seem inconsistent with regard to the quality of students' self-regulation: While results from other studies indicate insufficient self-regulation for complex tasks the results of this study indicate that students are well aware of the special demands of complex tasks and plan to use adequate learning strategies. One potential explanation for these inconsistencies concerns the different stages of learning: Students might be able to *plan* adequate self-regulation based on their adequate metacognitive knowledge about strategies (this study) but they might be unable to *enact* the planned strategies (due to cognitive overload or due to production or motivation deficits; other studies). To my knowledge only one other study explicitly focused on students' preparatory stages of learning: The corresponding exploratory study of this research project (Stahl, Pieschl, & Bromme, 2006). The results of this study are highly consistent with the exploratory study and seem even better as the calibration indices are partly even higher.

### ***3.3.3.2 The Impact of Prior Domain Knowledge and Epistemological Beliefs***

With regard to the impact of *internal conditions* on students' *discrimination* and *calibration*, students' prior domain knowledge (biology students vs. humanities students) as well as their

epistemological beliefs (captured by the scales *EBI-definitude*, *CAEB-texture*, and *CAEB-variability*) demonstrated significant impact.

#### *Prior Domain Knowledge*

Prior domain knowledge elicited a total of two effects: It was related to judgments on the COPEs-factor *deep processing*. An interaction effect indicates more fine-grained *discrimination* for biology students than for humanities students. Consistently, biology students were better *calibrated* with regard to their “estimated number of concepts”. Thus, in both cases prior knowledge was helpful at giving more differentiated judgments. Therefore, it can be concluded that prior knowledge might have helped students to perceive more fine-grained nuances of differences in tasks and thus might have facilitated giving their judgments accordingly. Students lacking adequate domain knowledge on the other hand might have based their judgments on surface cues of the task not considering the deep structure.

These effects are consistent with the predictions of the COPEs-model (Winne & Hadwin, 1998; chapter 2.1): Prior domain knowledge impacted students’ planned standards (captured by their “estimated number of concepts”) and their planned operations (captured by the COPEs-factor *deep processing*). Additionally, these results are mostly consistent with those of previous empirical studies about the impact of prior domain knowledge in the *preparatory* stages of learning (chapter 2.5). These studies demonstrated that prior domain knowledge had little quantitative impact but had some qualitative impact: Experts used more elaborate criteria to evaluate tasks and judged task difficulty more accurately (Chi, Feltovich, & Glaser, 1981; Chi, Glaser, & Rees, 1982; Lodewyk & Winne, 2005; Smith, 1990). Consistently, almost no (quantitative) effects were detected in this study and the only significant effects indicates that prior domain knowledge might enable students to differentiate between tasks of different complexity on a more fine-grained level. However, the results of this study do not consistently indicate that students with more prior domain knowledge were able to judge task complexity more accurately (Chi, Glaser, & Rees, 1982) as no effects for the “classification according to *Bloom-Categories*” was detected.

#### *Epistemological Beliefs*

Epistemological beliefs elicited four *discrimination* effects and one *calibration* effect. More specifically, main effects of *CAEB-variability* indicate that more “sophisticated” beliefs are consistently associated with judging *deep processing* and *dealing with multiple information sources* more important across all tasks and with classifying tasks in more complex *Bloom-Categories*. An additional interaction between *CAEB-variability* and task complexity for *dealing with multiple information sources* indicates that these effects become less pronounced for more complex tasks. Most likely, students who believe that knowledge is variable and dynamic automatically consider all kinds of tasks more complex per se (effect on “classification according to the *Bloom-Categories*”). In order to counteract this perceived complexity and in order

to adequately deal with the perceived variability of knowledge they might plan strategies of deeper elaboration (effects on *deep processing* and *dealing with multiple information sources*). It could be said that these “sophisticated” students discriminate between tasks on a higher level. However, the *calibration* effect is counterintuitive: The belief in definite knowledge (*EBI-definitude*) was positively associated with better calibration with regard to “estimated number of concepts”: Students with more “naïve” epistemological beliefs appear to be better at adapting their estimates to task complexity while “sophisticated” students showed less flexibility in their judgments. One potential explanation related to the measurement of epistemological beliefs is feasible: The scale *EBI-definitude* reaches from views that knowledge is definite (“naïve” *absolutist*) to views that knowledge is indefinite (“sophisticated” *relativist*) but does not capture most “sophisticated” flexible *evaluativist* epistemologies (Kuhn, Cheney, and Weinstock, 2000). Most likely, students with such epistemological beliefs would give judgments near the scale mean for *EBI-definitude*. The frequency distribution for *EBI-definitude* reveals, that judgments in this sample range from very indefinite conceptualizations (*relativist*) to moderately definite ones (probably *evaluativist*); no very definite judgments were given. Students with moderately definite views on *EBI-definitude* – probably the most “sophisticated” students according to this proposed explanation – possess higher *calibration* indices than students who considered knowledge very indefinite.

These effects are consistent with the predictions of the COPEs-model (Winne & Hadwin, 1998; chapter 2.1): Epistemological beliefs impacted students’ planned standards (captured by their “estimated time”), their planned operations (captured by the COPEs-factors *deep processing* and *dealing with multiple information sources*), and their evaluations (captured by their “classification according to the *Bloom-Categories*”). Additionally, these results are mostly consistent with those of previous empirical studies (chapter 2.6). So far, most studies support the notion of a beneficial main effect of “sophisticated” epistemological beliefs consistent with the main effects within the *discrimination* analyses (Bartholomé, Stahl, Pieschl, & Bromme, 2006; Bråten & Strømsø, 2004; Kardash & Scholes, 1996; Mason & Boscolo, 2004; Mason & Scirica, 2006; Ryan, 1984; Schommer, 1990; Schommer, Crouse, & Rhodes, 1992; Schommer-Aikins & Hutter, 2002). For example, in other studies concentrating on the *preparatory* stages of learning students with “sophisticated” beliefs perceived the affordances of ill-structured tasks more accurately (King & Kitchener, 2002) and set more adequate goals (Bråten & Strømsø, 2004; Ryan, 1984). However, the interaction effect of the *discrimination* analyses and the *calibration* effect (see above) are inconsistent with theoretical expectations and empirical results: Theoretically it was assumed that more “sophisticated” students should display more flexible adaptations (Hammer & Elby, 2002). Consistently, empirical studies investigating the relationship between students’ epistemological beliefs and their calibration in the traditional sense (chapter 2.2) found that “sophisticated” beliefs in gradual learning (*quick learning*, Schommer, 1990) as well as in complex knowledge (*simple knowledge*, Schommer, Crouse, & Rhodes, 1992) were associated with less overestimation of

comprehension. Furthermore, an exploratory study for this study investigating the same questions with similar methodology (Stahl, Pieschl, & Bromme, 2006) demonstrated that “sophisticated” beliefs were associated with better *calibration* indices in the preparatory stages of learning. To conclude: These latter two effects are inconsistent with theoretical assumptions as well as with previous empirical results and remain enigmatic.

### 3.3.3.3 *Open Issues and Implications*

This study was successful at demonstrating that students in fact do *discriminate* between tasks of different complexity and *calibrate to task complexity* and that these processes are impacted by students’ *prior domain knowledge* and *their epistemological beliefs* in the *preparatory* stages of learning (previous sections). However, some limitations have to be considered.

#### *Open Issues and Limitations*

In this section most involved variables will be scrutinized with regard to their validity, corresponding limitations, and open issues.

First of all, the *COPEs-questionnaire* will be discussed: Based on an exploratory study (Stahl, Pieschl, & Bromme, 2006) a subset of items was pre-selected as dependent variables for this study. Consider that it is not surprising that only a subset of COPEs-items is suited to capture students’ *calibration*. The whole questionnaire was developed to capture the *preparatory* stages of self-regulated learning most thoroughly. Thus, additional patterns of answers are feasible: For example, some items might capture students’ general trait-like approaches to learning. Such lack of systematic relationships between students’ judgments and *task complexity* was found in the exploratory study for a subset of items termed “no trend”. Or students might adapt their answers to other external conditions than task complexity, for example whether it required application or not (see items with “specific” trends for *apply* in the exploratory study). Therefore, all items were retained because the additional patterns are also interesting, meaningful and contribute to a more thorough understanding of the *preparatory* stages of learning even though they do not capture *calibration*.

A second open issue concerns the *absolute* quality of students’ calibration. Even though students in general are quite successful at discriminating between and calibrating to tasks of different complexity they might be still far from perfect. Overestimating the complexity of simple tasks might not be detrimental for learning, just not be the most parsimonious way to solve these simple task. Imagine for example a student who considers a simple factual multiple-choice task more complex than necessary. Misjudging the complexity of more complex tasks on the other hand might have more detrimental effects. Not only would the answer be less adequate, but also the gained understanding would be more superficial than required. Data from this study (as well as from the corresponding exploratory study; Stahl, Pieschl, & Bromme, 2006) tentatively indicates that students might in fact underestimate



the complexity of very complex tasks. For example, the calibration graph depicting students' *absolute* calibration for the item "classification according to the *Bloom-Categories*" indicates that students classify complex tasks (*analyze – create*) into less complex *Bloom-Categories* than indicated. Furthermore, the calibration graphs from all other items and COPEs-factors also indicate that students' average judgments are never even close to the extremes of the scales. Thus, students could – or probably should – consider indicators of deep processing even more important for complex tasks than they currently do.

A third issue pertains to *prior domain knowledge*. This learner characteristic had surprisingly little impact on students' task definitions, goals and plans captured by the COPEs-questionnaire. This is clearly inconsistent with the huge impact of prior domain knowledge detected for later stages of learning (*enactment*, chapter 2.5). A potential explanation concerns the domain-specificity versus domain-generality of expertise: Students' task definitions, goals and plans might be more dependent on domain-general approaches to learning (students' metacognitive knowledge about tasks and adequate strategies) than on prior domain knowledge for a specific topic thus explaining the minimal effects.

A related fourth open issue pertains to the question to which degree *epistemological beliefs* are domain related or independent from the domain they refer to. There is growing evidence that learners have general epistemological beliefs, as well as domain related beliefs (Buehl, Alexander & Murphy, 2002). But up to now it is unclear how such different levels might interact with each other. Due to these discussions one domain-general instrument (EBI) and one domain-dependent instrument (CAEB) were included in this study. In this study focusing on the preparatory stages of learning, the domain-dependent epistemological beliefs of students were more important than their domain-general beliefs. Furthermore, the scale *CAEB-variability* elicited most effects. This is consistent with the fact that this epistemological dimension also elicited the most consistent effects in previous studies (chapter 2.6): The "sophisticated" belief in relative and variable knowledge was consistently associated with reporting more deep elaboration strategies and with reporting less surface strategies (Bråten & Strømsø, 2005, 2006; Cano, 2005; Dahl, Bals, & Turi, 2005).

### *Implications*

Given the significant effects in these *preparatory* stages of self-regulated learning, these stages should receive more attention not only in research but also in educational settings. To give an example: Most computer-assisted learning environments (CALs) include different kind of help functions. Studies show that especially those learners who would need this help the most do not use it appropriately (Aleven, Stahl, Schworm, Fischer, Wallace, 2003). Inappropriate task definitions and goal setting and planning might be one reason for such findings (e.g., underestimation of task complexity). Students should fail, if their premises about tasks and goals are wrong. Based on two assumptions a tentative implication can be suggested: First, even though students are good self-regulated learners in these preparatory

stages of learning they are far from perfect (see above). And second, epistemological beliefs have significant impact on students' task definitions, goals, and plans (see above). Consequently, it should be possible to devise an instructional intervention focusing on epistemological issues and therefore fostering even better self-regulated learning processes in this preparatory stage of learning. It might be helpful to explicitly instruct students about task demands and the epistemology of a domain. Even though, this kind of scaffold failed its purpose in this study, a revised version might prove successful in subsequent studies.

### 3.4 Study II – Calibration to Task Complexity in the Enactment Stages

This study pursued the following research questions: (1) Do students *discriminate* between *tasks* of different complexity (different *Bloom-Categories*)? (2) Do students *calibrate* to *task complexity*? (3) Do the *internal conditions* prior domain knowledge and epistemological beliefs impact these processes of *discrimination* and *calibration*? (4) What factors determine the learning outcome (e.g., students' *calibrations* or *internal conditions*)?

Biology students ( $n = 14$ ) and humanities students ( $n = 21$ ) who received either an epistemological sensitization that elicited more “sophisticated” epistemological beliefs or a neutral introduction to the domain solved five tasks of different *Bloom-Categories* with a hypertext on “genetic fingerprinting”. During their learning process, logfiles and their concurrent thoughts were captured. Subsequently, students were interviewed. Results indicate that students *discriminate* between tasks of different complexity and *calibrate* their learning process to task complexity (e.g., by accessing fewer hypertext nodes for simpler tasks than for complex tasks). Furthermore, these discrimination and calibration processes were significantly related to their prior domain knowledge and to their epistemological beliefs (*internal conditions*). Humanities students spent more time on tasks and reported using more “planning” strategies but were outperformed by biology students (prior domain knowledge). And students with the epistemological sensitization spent more time and reported using more “planning” strategies for the complex *evaluate* task.

#### 3.4.1 Method

##### 3.4.1.1 Procedure

All students received 18 € reimbursement for participation. To investigate the impact of prior domain knowledge, students were selectively recruited to ensure two levels of knowl-

edge: Biology students ( $n = 14$ ) were recruited during regular courses in biology, humanities students ( $n = 21$ ) were recruited by a posting at the psychological institute.

During the first online session, students filled in two questionnaires about their domain-dependent epistemological beliefs (CAEB and GCBS), which took them about 15 minutes (all materials from the online session can be found in appendices B1 – B2). Twenty biology students and 22 humanities students completed the online-questionnaires. The second session encompassed an individual face-to-face session and lasted approximately two hours (all materials from the face-to-face session can be found in appendices B5 – B20). Not all students continued: 14 biology students (70 % of the original sample) and 21 psychology students (95 % of the original sample) participated in this second session. First, students had to answer a short molecular genetics knowledge test. Then, an epistemological sensitization was administered: Students were sorted into two matched sub-samples, one sub-sample received a neutral introduction to genetic fingerprinting, and the other sub-sample received an epistemological introduction in order to elicit a more “sophisticated” view (the same factual information enriched with comments about the epistemological nature of the presented facts). Subsequently, the epistemological beliefs instruments (CAEB and GCBS) were re-administered as a treatment check of this intervention. Then all students were introduced to the structure and navigational options of the hypertext. In the main part of this session, students solved 3 – 5 tasks representing different *Bloom-Categories* of different complexity (in temporal order: *remember*, *remember*, *evaluate*, *understand*, and *remember*) with the hypertext on “genetic fingerprinting”. For all tasks, students had to write down their answers. During this approximately one hour long learning phase, students were prompted in fixed time intervals to elaborate on their concurrent thoughts. Furthermore, detailed logfiles were automatically collected to capture students’ concurrent navigational actions in the hypertext. At the end, two retrospective stimulated recall interviews were conducted (for a simple *remember* task and for the complex *evaluate* task).

In all analyses regarding the impact of *internal conditions* prior domain knowledge was included as factor (biology students vs. humanities students) and the epistemological sensitization (neutral introduction vs. epistemological introduction) was also included as a factor.

### 3.4.1.2 Participants

Although the advanced students of biology ( $n = 14$ ) were no experts in the specific topic of genetic fingerprinting they can be considered discipline experts (chapter 2.5.3). These biology students (4 males, 10 females) were on average 24 years old ( $SD = 2.46$ ) and studied in the 9<sup>th</sup> semester biology or related majors ( $SD = 2.24$ ). They already attended 7 of 10 relevant courses in molecular biology. Adequate background knowledge in molecular biology was verified by the results of a short knowledge test ( $M = 7.57$ ,  $SD = .65$ , with 8 points maximum; the knowledge test can be found in appendix B5). Their interest ( $M = 4.50$ ,

$SD = .76$ , on a 5-point scale with 1 = very low and 5 = very high) and self-rated prior domain knowledge in molecular biology ( $M = 3.86$ ,  $SD = 0.66$ , on a 5-point scale with 1 = very low and 5 = very high) were also quite high.

Students of humanities ( $n = 21$ ) can be considered novices (chapter 2.5.3). These humanities students (1 male, 20 females) were on average 21 years old ( $SD = 1.34$ ) and studied in the 3<sup>rd</sup> semester ( $SD = 4.00$ ) a humanity major such as psychology, history or sociology. They did not attend any of the 10 relevant university courses in molecular biology. Low background knowledge in molecular biology was verified by the results of a short knowledge test ( $M = 2.52$ ,  $SD = 1.78$ , with 8 points maximum). Their interest ( $M = 2.62$ ,  $SD = 0.74$ , on a 5-point scale with 1 = very low and 5 = very high) and self-rated prior domain knowledge in molecular biology ( $M = 1.62$ ,  $SD = 0.67$ , on a 5-point scale with 1 = very low and 5 = very high) were also quite low.

The difference between biology students and humanities students with regard to their prior domain knowledge was significant on all relevant variables (points in the molecular biology test:  $t(33) = 10.14$ ,  $p < .001$ ; number of attended courses relevant to molecular biology:  $t(32) = 13.43$ ,  $p < .001$ ; self-rated prior domain knowledge:  $t(33) = 9.73$ ,  $p < .001$ ). In all instances, biology students displayed more prior domain knowledge. Thus, prior domain knowledge indicated by these quasi-experimental groups (biology students vs. humanities students) will be used as dichotomous independent variable.

### 3.4.1.3 Materials

#### *Epistemological Beliefs Questionnaires*

A combination of two domain-dependent instruments that had to be completed in reference to the domain of genetics was used to measure epistemological beliefs. The CAEB (Connotative Aspects of Epistemological Beliefs; Stahl & Bromme, in press; appendices B2 and B8) includes 24 pairs of connotative adjectives to measure the dimensions *texture* (structure and accuracy of knowledge; sample item: “structured – unstructured”) and *variability* (stability and dynamics of knowledge; sample item: “dynamic – static”). The GCBS (Global Certainty Beliefs Scale, Trautwein & Lüdke, in press; appendices B3 and B9) captures more declarative aspects of beliefs about the *certainty* and attainability of scientific knowledge with 7 items (sample item: “Scientific laws are universal truths.”). All statements and adjectives in both instruments were rated on 7-point scales.

Because of the small sample size ( $n = 35$ ), no exploratory factor analyses were computed. Instead, the factor solutions suggested by the original publications were scrutinized with regard to their reliability within this sample. Both instruments were administered twice, once before and once after the epistemological sensitization. The original scale *CAEB-texture* comprises 10 items and proved satisfactory reliability in this sample pre-instructionally (Cronbach’s  $\alpha = .84$ ) as well as post-instructionally (Cronbach’s  $\alpha = .85$ ).

The original scale *CAEB-variability* comprises 7 items and proved satisfactory reliability in this sample pre-instructionally (Cronbach's  $\alpha = .74$ ) as well as post-instructionally (Cronbach's  $\alpha = .90$ ). The 7-item GCBS scale was labeled *GCBS-certainty* and proved satisfactory reliability pre-instructionally (Cronbach's  $\alpha = .72$ ) as well as post-instructionally (Cronbach's  $\alpha = .83$ ). In this study, these epistemological beliefs scales were only used to determine the effectiveness of the epistemological sensitization. Because it was successful, these scales will not be included in any further analyses.

### *Epistemological Sensitization*

After the first test of the epistemological sensitization only proved marginally successful (study I), the instructions were revised. Again, this epistemological sensitization served two aims: First, the epistemological introduction was intended to elicit more “sophisticated” epistemological beliefs whereas the neutral introduction should not systematically change students' epistemological beliefs. Second, both introductions should provide all students with a minimum of basic knowledge in molecular genetics. Thus, they adequately contextualized students to the topic of genetic fingerprinting.

In this study, a revised version of the epistemological sensitization was tested (the German versions of these introductions can be found in appendices B6 and B7). The most significant changes pertain on the one hand to the length of the introductions (they were shortened) and now only include three parts: (1) definition of genetic fingerprinting, (2) biological background about DNA (structure of DNA; coding and non-coding DNA regions; DNA loci and alleles; mutation of DNA), and (3) biological background about human genomes (nuclear and mitochondrial). On the other hand, the density of epistemological comments was enhanced by inserting more references to scientific controversies. In total, the revised epistemological introduction contains 55 comments about epistemological views. As in the first prototype, these comments stress the *relativistic* as well as the *absolutistic* view and thus serve to sensitize students with regard to such epistemological questions. This introduction was meant to elicit an *evaluativistic* view. The revised neutral introduction encompasses 8 pages with 1.291 words. The revised epistemological introduction encompasses 8 pages with 1.625 words. For an example of the implemented changes see Table 8.

Empirical results show that the two sub-samples receiving the two versions of the introduction were adequately matched with regard to their epistemological beliefs (no significant pre-instructional differences could be detected: *CAEB-texture*:  $F(1,33) = 1.38, p = .248$ ; *CAEB-variability*:  $F(1,33) = .01, p = .937$ ; *GCBS-certainty*:  $F(1,33) < .01, p = .965$ ). Additionally, consistent effects of the epistemological sensitization could be detected in a repeated-measure analysis: A significant multivariate interaction ( $F(1,31) = 4.73, p = .008$ ) was replicated univariately on the two CAEB scales (also Figure 3.4-1; *CAEB-texture*:  $F(1,33) = 13.02, p = .001$ ; *CAEB-variability*:  $F(1,33) = 7.95, p = .008$ ). For *GCBS-certainty*, this interaction was not significant ( $F(1,33) = .33, p = .572$ ). Furthermore, significant uni-

variate main effects for the repeated-measure factor emerged for the two CAEB scales (also Figure 3.4-1; *CAEB-texture*:  $F(1,33) = 6.41, p = .016$ ; *CAEB-variability*:  $F(1,33) = 6.81, p = .014$ ). The main effects indicate that all students became more “sophisticated” after reading the revised introductions. The interactions consistently demonstrate that students who read the epistemological introduction became significantly more “sophisticated” than the students who read the neutral introduction. The effects point in the same direction for *GCBS-certainty* but were not significant.

Table 8: Excerpts from the epistemological introductions, epistemological comments are highlighted in italics, additional comments in the revised introduction are underlined

Epistemological introduction – study I	Epistemological introduction – study II
<p><i>According to the present state of knowledge, a gene can be considered as the basic unit of genetic information. It involves a fragment of the DNA-sequence that contains all information for building proteins. As a gene thus always has a function – although a mostly unknown one – DNA-fragments that contain genes are called coding DNA regions. The different conditions of a gene are labeled alleles. For example, a gene for the color of a flower could possess three alleles, red, white and rose. Human genes do not usually function in such a simple way. On the contrary, complex interactions between multiple genes – partly also involving none-coding regions – determine human characteristics.</i></p>	<p><i>According to the present state of knowledge, a gene can be considered as the basic unit of genetic information. <u>It is worth considering though, that some scientists presently plead for a re-definition of the concept “gene” as current research results are not consistent with the traditional definition. Nonetheless, this introduction uses the traditional definition according to which a gene can be considered a fragment of the</u> DNA-sequence that contains all information for building proteins. As a gene thus always has a function – although a mostly unknown one – DNA-fragments that contain genes are called coding DNA regions. <u>The separation of coding and none-coding DNA regions has not yet been achieved conclusively, though. One of the reasons is the ongoing discovery of new functions of DNA regions.</u></i></p> <p>The different conditions of a gene are labeled alleles. For example, a gene for the color of a flower could possess three alleles, red, white and rose. <i>Human genes do not usually function in such a simple way. On the contrary, complex interactions between multiple genes – partly also involving none-coding regions – determine human characteristics.</i></p>

To summarize, this second empirical test of the epistemological sensitization can be considered moderately successful: The results consistently point in the intended direction (post-instructionally students with the epistemological introduction possessed more “sophisticated” beliefs than students with the neutral introduction) and the epistemological introduction was successful to elicit more “sophisticated” epistemological beliefs for the majority of epistemological beliefs scales. Only for *GCBS-certainty* these effects were not significant, either because of a ceiling effect (even before the epistemological sensitization students’ beliefs were very “sophisticated”) or because the denotative statements used in

this instruments might not be as well-suited to capture minute changes in epistemological beliefs as the connotative adjective pairs of the CAEB. Because of the moderate success of the epistemological sensitization the subsequent statistical analyses *will* consider these experimental groups (neutral introduction vs. epistemological introduction). Additionally, the introductions can be considered beneficial because they also served their second purpose well: They contextualized students adequately to the topic of “genetic fingerprinting”.

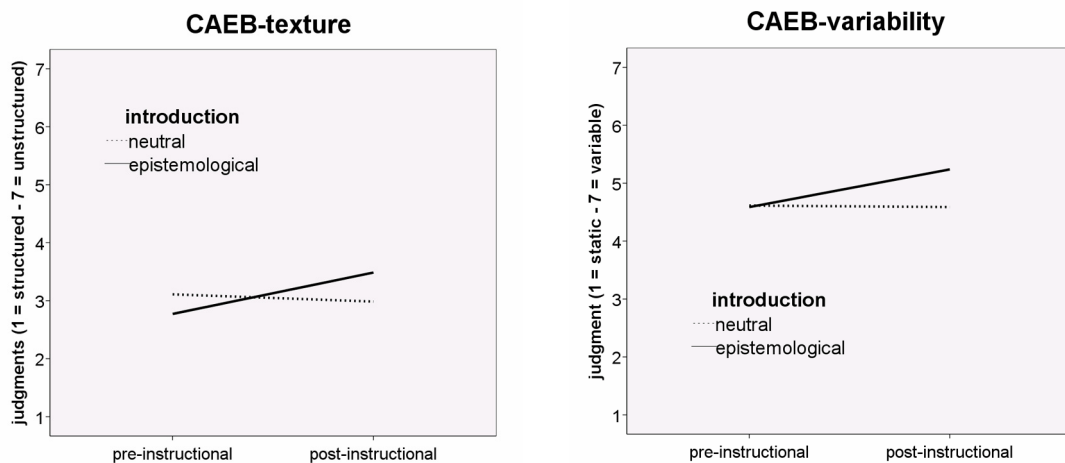


Figure 3.4-1: Students' epistemological beliefs on the scales *CAEB-texture* (left) and *CAEB-variability* (right) as a function of the epistemological sensitization (neutral introduction vs. epistemological introduction).

### *Tasks of Different Complexity*

Bloom's revised taxonomy (Anderson et al., 2001; chapter 2.3.2) distinguishes between task classes affording cognitive processes of different complexity (in order of ascending complexity): (1) *remember*, (2) *understand*, (3) *apply*, (4) *analyze*, (5) *evaluate*, and (6) *create*. For this study, it was impossible to retain the tasks from study I as several constraints had to be fulfilled. Thus, five new tasks were developed.

First, students in study I had to *evaluate* tasks from all *Bloom-Categories*. In this study on the other hand, students had to *solve* tasks with the hypertext on “genetic fingerprinting”. Due to time constraints it was impossible to cover all six *Bloom-Categories*. Second, the tasks had to be answerable with the help of the hypertext. This was not the case for all tasks used in study I. Thus, some tasks could not be re-administered but were used as templates for the development of comparable new tasks. Third, all tasks should require students to refer to the hypertext and not be answerable without it. Thus, no task of the very complex but abstract *Bloom-Category create* was included. Fourth, the order of tasks should require the students on the one hand to enhance their depth of processing (simple tasks should be followed by complex ones) and on the other hand it should also require students to diminish their depth of processing (complex tasks should be followed by simple ones). Only such an active regulation of processing can be interpreted as a clear sign of students' adaptation to

task complexity. If students just continually enhanced their depth of processing this could rather be due to the fact that they got more familiar with the hypertext. All new tasks were classified to *Bloom-Categories* by an expert panel.

Resulting, the following tasks were administered in a fixed order: First, one very simple *remember* task was used as an ice-breaker to familiarize students with the hypertext, followed by a second *remember* task (“When analyzing a multi-copy Y-STR-locus ... (1) ... *multiple alleles can be diagnosed per individual.* (2) ... the genetic material has to come from different individuals. (3) ... only the first motives have to be analyzed per individual. (4) ... the result indicates contamination. (5) ... only one allele can be diagnosed per individual.” (correct answer in italics)). To investigate students’ enhancement of processing with ascending task complexity, a complex *evaluate* task was utilized next (Figure 3.4-2). This task requires critical evaluation of the different methods of genetic fingerprinting with regard to paternity testing and thus affords reading and integrating content from multiple hypertext nodes. To create a most natural learning scenario, students were allowed to work through the tasks in a self-paced way. But all students were at least required to solve the first three tasks. To further scrutinize these processes, in-depth stimulated recall interviews were conducted for the second *remember* task and the third *evaluate* task (see below). The subsequent tasks on the other hand required a diminishing level of elaboration and were optional: An *understand* task was administered that was only slightly more complex than the *remember* tasks but had to be answered in an open format. And last, another *remember* task was used. All utilized tasks can be found in appendix B17.

---

### **Paternity Testing**

Imagine you are a molecular biologist and got a request from an employee of Pro Familia (a German organization involved in counseling with regard to family planning issues).

More specifically, this employee describes that Pro Familia wants to offer counseling with regard to the conflicts surrounding the issue of biological paternity. Thus, this employee wants to understand the biological background of paternity testing. He reports to have heard that the analysis of biological paternity always involves Y-STR analysis. Considering this background, he wants the following questions answered:

- Does it make sense to use Y-STR analysis?
- Or are there other methods?
- What kind of method would you recommend? Why?

It is your task to answer the questions of the Pro Familia employee. Please write down a consistent answer that considers all his questions.

---

Figure 3.4-2: The complex *evaluate* task that was used.



In order to determine the learning outcome, students' answers to these tasks were graded. All *remember* tasks used the multiple-choice format, thus the correctness of students' answers could be determined easily. The *evaluate* and *understand* tasks on the other hand involved open answers. Thus, the number of written words was counted for these two tasks. Furthermore, these more complex tasks were scored qualitatively with two task-specific rubrics. The scoring rubric for the *evaluate* task contains four sub-scores in total. The first score concerns the number of mentioned DNA analysis methods (0 – 4 points can be given for mtDNA analysis, STR analysis, Y-STR analysis, and RFLP analysis). The second score concerns the elaborateness of the description of each DNA analysis method (0 – 2 points can be given for each method). Third, the correctness of the students' conclusions with regard to each DNA analysis method is evaluated: One point is given if mtDNA analysis is excluded because the mitochondria are inherited maternally, another point is given if the limitations of Y-STR analysis are diagnosed correctly (this analysis can only be conducted for male children), and two points are given if STR analysis is considered the best method. Further extra points (0 – 2) can be obtained if a good conclusion is written. From these scores, also the correctness can be deducted (recommendation of STR analysis). The scoring rubric for the *understand* task on the other hand only contains two dimensions: number of mentioned arguments and correctness of the given explanation.

To conclude, the subsequent statistical analyses will include task complexity as defined by these task representing different *Bloom-Categories* as independent variable. Note that the tasks will not be re-ordered according to ascending complexity, but will be *retained in their temporal order* to explicitly visualize students' ascending as well as descending depth of processing (see above: *remember*, *remember*, *evaluate*, *understand*, and *remember*). Additionally, students' answers will be used as dependent variables capturing their *learning outcome*. For each task *task difficulty* can be determined (percentage of correct answers) and for the tasks *evaluate* and *understand* the number of written words can be considered as well as the sub-scores of the scoring rubrics.

#### *Logfiles Capturing Students' Concurrent Actions*

During students' task solution all individual concurrent navigational actions within the hypertext on "genetic fingerprinting" were recorded (chapter 3.2). From these logfiles, two scores were computed that were used as repeated-measure dependent variables in all subsequent analyses: *Time for task completion* (TTC) contains the exact time a student needed to complete each single task. And the *number of accessed nodes* (NAN) explicates how many nodes a student accessed for each task.

#### *Prompts Eliciting Students' Concurrent Thoughts*

Students' concurrent thoughts were captured by asking them "What are you currently thinking about?" in fixed time intervals (approximately every two minutes). This method-

ology was chosen instead of thinking aloud to capture students' concurrent thoughts because a pilot study with the same hypertext (Kulbe, 2006) indicated that utterances during thinking aloud hardly ever referred to students' metacognitive processes or their epistemological views. This more invasive methodology of prompting (asking direct questions) on the other hand should trigger students' deeper explanations. Ericsson and Simon (1980) outlined the differences between such prompting for reflection and the thinking aloud procedure. According to this cognitive model, only the conscious content of short term memory (STM) can be verbalized. The content of long term memory (LTM), however, can not be verbalized directly, but first has to be retrieved to STM. Ericsson and Simon (1980) distinguish three levels of verbalization. In *level-1 verbalization*, the participant vocalizes her verbally encoded content of STR (talk aloud). In *level-2 verbalization*, the participant employs some mediation processes to encode all content of STM verbally or explicate it. These mediation process require processing time but are not assumed to affect the structure of the cognitive processes (think aloud). In *level-3 verbalization*, the participant has to explain her thoughts, ideas or motives. These interpretative processes not only require processing time but also may change the cognitive processes because they require information to be retrieved from LTM and linked with the content of STM. Bannert (2003) refers to these level-3 verbalizations as reflection prompts. The prompting employed in this study is analogous to this idea.

In order to systematically analyze students' answers to these prompts, a coding scheme for the verbalizations of their concurrent thoughts was developed (the coding scheme can be found in appendix B19). Each answer was sorted into one of three superordinate categories: (1) "planning" (PL), (2) "enactment" (EN) or (3) "reflection / revision" (REV). Each of these categories encompasses multiple sub-categories. For example, the category "planning" contains the sub-categories "planning of task solution", "planning of time", "planning of navigation", and "planning of writing the answer". Two raters blindly coded the protocols of a sub-sample of 12 students (34 % of the total sample). For 73.4 % of the prompts ( $n = 259$ ) the two raters assigned the same superordinate category, for 66.2 % of the prompts ( $n = 229$ ) they assigned not only the same superordinate category, but also the same sub-category. All differences were resolved by discussion. Subsequently, one of the two raters coded the remaining protocols.

From these concurrent thoughts, three scores were computed that were used as repeated-measure dependent variables in all subsequent analyses. The number of concurrent thoughts categorized as "planning" (PL), "enactment" (EN), and "reflection / revision" (REV) were determined for all tasks. The subcategories were not further analyzed.

#### *Retrospective Stimulated Recall Interviews*

To validate students' concurrently verbalized thoughts and actions (see previous paragraphs), an in depth retrospective stimulated recall interview was conducted twice, once for

the second *remember* task and once for the complex *evaluate* task. In both cases, students were asked in detail about all stages of their self-regulated learning process (their *task definition, goal setting and planning, enactment, and potential adaptations*).

In each standardized interview, the same questions were posed (the experimenter guide for the interview can be found in appendix B18). Note though that interviewers could skip questions if students already gave relevant answers. First, in order to remind students of their actions during task solution, students were presented or “stimulated” with their own “history” of their navigational actions in the hypertext (logfiles). If they wanted, they could re-visit some nodes to refresh their memory. Each interview started with describing two approaches to task solution. For example, for the simple *remember* tasks one described approach mainly focused on a fast task solution, which comprised comparison of multiple-choice alternatives with sentences in specific hypertext nodes. The second described approach mainly focused on deeply understanding the underlying content by reading and comparing multiple hypertext nodes. Subsequently, the students had to elaborate their task understanding and could do so by referring to those two approaches (*task definition* in the COPEs-model). Afterwards, students had to describe their initial goals (*goal setting and planning*) and describe their task solution process in detail (*enactment*). Furthermore, they were explicitly asked if they changed their strategy during task completion (*adaptation*).

In order to systematically analyze students’ answers in the interview, a complex coding scheme was developed based on students’ utterances (the coding scheme can be found in appendix B20). For each stage of the COPEs-model, students’ answers were assessed on multiple dimensions. To give an example, consider the coding scheme for the *task definition* stage: First, it is noted whether students’ more strongly agreed with the first described approach or with the second one. Second, students’ evaluation of *task complexity* is coded on a three point scale. Third, the main argument for simplicity or complexity is coded as well as the number of reported arguments. Furthermore, the raters evaluated whether the approach described by the individual students matched rather the first or second described approach. Similar dimensions exist for the stages of *goal setting and planning* and *enactment*. Two raters separately coded the protocols for a sub-sample of 21 students (60 % of the total sample). For 75.3 % of the categories and evaluations ( $n = 1386$ ) the two raters agreed. All differences were resolved by discussion. Subsequently, both raters jointly coded the remaining interviews.

From these retrospective stimulated recall interviews, multiple variables were derived capturing students’ whole self-regulated learning process, but only for one *remember* and the *evaluate* task. These variables were utilized as additional qualitative dependent variables.

### 3.4.2 Results

In this study  $p < .05$  was defined as significant and  $p < .10$  as marginally significant. Because most students ( $n = 29$ , 82.8 % of the sample) finished all five tasks, statistical analyses will be performed with task complexity defined by all five tasks of different *Bloom-Categories* as a repeated-measure factor (in temporal order: *remember*, *remember*, *evaluate*, *understand*, and *remember*). Variables capturing students' concurrent actions (logfiles: TTC and NAN) and variables capturing students' concurrent thoughts (their answers to the prompts categorized as: PL, EN, and REV) were considered as dependent learning *process* variables. Additionally, students' written answers were analyzed as dependent learning *outcome* variables (task difficulty, number of written words, sub-scores of the qualitative scoring rubric). The data from the retrospective stimulated recall interview and more detailed logfiles were scrutinized as additional qualitative data source.

#### 3.4.2.1 Do Students Discriminate Between Tasks of Different Complexity?

To answer the first research question, within-subject repeated-measure analyses across all five tasks of different *Bloom-Categories* (in temporal order: *remember*, *remember*, *evaluate*, *understand*, and *remember*) were calculated for all dependent variables. To enrich these quantitative analyses with more qualitative data, the stimulated recall interview and detailed logfiles will be scrutinized for the second *remember* and for the *evaluate* task.

##### *Quantitative Analyses*

A within-subject repeated-measure MANOVA was calculated across all five tasks for students' *concurrent actions* captured by logfiles (TTC = *time for task completion* and NAN = *number of accessed nodes*). Results indicate a significant multivariate main effect for the repeated-measure factor task ( $F(8,20) = 10.20, p < .001$ ) that was replicated univariately on both dependent variables (NAN:  $F(4,108) = 43.67, p < .001$ ; TTC:  $F(4,108) = 32.11, p < .001$ ). The corresponding *calibration* graphs show students' enhanced processing for more complex tasks as well as their diminishing processing for simpler tasks. For both dependent variables the calibration graphs indicate a "peak" for the complex *evaluate* task. While students spent on average between five and ten minutes on the simpler tasks, they spent on average approximately 20 minutes on the complex *evaluate* task (*time for task completion*, TTC). The picture for *number of accessed nodes* (NAN) is similar as can be seen in Figure 3.4-3.

A within-subject repeated-measure MANOVA was calculated across all five tasks for students' *concurrent thoughts* captured by their answers to prompts (PL = "planning", EN = "enactment", and REV = "reflection / revision"). Results indicate a significant multivariate main effect of the repeated-measure factor task ( $F(12,23) = 27.55, p < .001$ ) that was replicated univariately on all dependent variables (PL:  $F(4,136) = 24.47, p < .001$ ; EN:

$F(4,136) = 30.46, p < .001$ ; REV:  $F(4,136) = 25.53, p < .001$ ). For all three types of concurrent thoughts, the overall picture was similar: Students gave more answers for the complex *evaluate* task than for the simpler tasks, indicated by a “peak” in the *calibration* graphs. For example, students gave on average less than one “reflection / revision” answer for the simple *remember* tasks, but on average they gave approximately 1,5 “reflection / revision” answers for the *understand* task and more than three for the complex *evaluate* task.

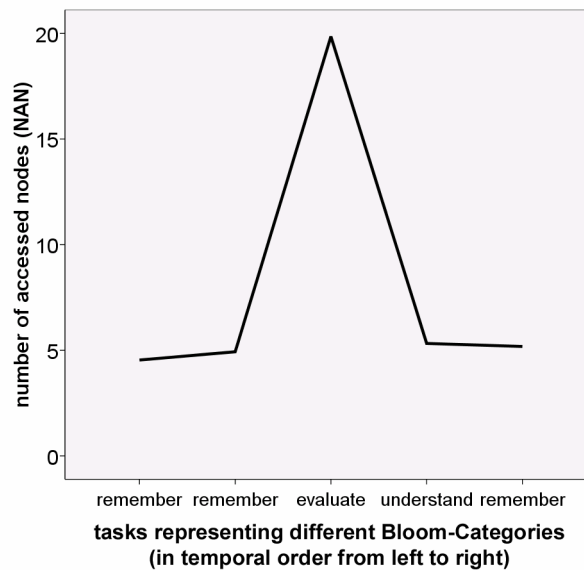


Figure 3.4-3: Calibration graph depicting students’ the number of accessed notes (NAN) as a function of task complexity (*Bloom-Categories*). The “peak” indicates a significant increase of accessed nodes for the complex *evaluate* task (enhanced processing) and a decrease of accessed nodes for the subsequent simpler tasks (diminished processing).

A t-test for dependent samples was calculated to compare the number of written words as one indicator of students’ *learning outcome* between the two open answer tasks. Results indicate that students wrote significantly more for the complex *evaluate* task (120 words) than for the simpler *understand* task (60 words;  $t(31) = 5.89, p < .001$ ). Additionally, a within-subject repeated-measure ANOVA was calculated across all five tasks for task difficulty (percentage of correct answers) as an additional indicator of students’ learning outcome. Results indicate no significant effect of the repeated-measure factor task ( $F(4,25) = 1.92, p = .138$ ). All tasks were solved correctly by more than 60 % of the students.

#### *Qualitative Insights from the Stimulated Recall Interview*

Students’ *perceptions* of their own *concurrent learning processes* were captured by their answers in the retrospective stimulated recall interview. Results indicate that students’ *task definitions*

according to the COPES-model differed significantly between tasks: For both tasks, one superficial approach and one deep approach were presented by the experimenter. All students who directly referred to these approaches indicated that they perceived the superficial approach to fit the *remember* task ( $n = 18$ ). For the *evaluate* task on the other hand most students indicated that the deep approach was the better fit ( $n = 14$ ), opposed to a minority of students ( $n = 4$ ) who indicated that they considered the superficial approach a better fit for the *evaluate* task. Thus, the described approaches differed significantly ( $X^2 = 22.91$ ,  $df = 1$ ,  $p < .001$ ). Furthermore, students' comments about task complexity were mapped on a 3-point scale. Students judged the *remember* task to be significantly simpler than the *evaluate* task ( $t(32) = -4.70$ ,  $p < .001$ ). Additionally, students chose significantly different arguments to justify their perception of task complexity ( $X^2 = 27.21$ ,  $df = 8$ ,  $p = .001$ ): For the *remember* task, the perceived simplicity was mostly justified by referring to the task format (multiple-choice,  $n = 8$ ) or the short time needed to find the correct answer ( $n = 7$ ). For the *evaluate* task on the other hand, the perceived complexity was also justified by referring to the task format (open answer,  $n = 5$ ), but also by referring to deeper criteria like the amount of information necessary to solve the task (many hypertext nodes needed to be accessed,  $n = 7$ ) or to the deep understanding necessary for task completion ( $n = 5$ ).

Students' *goal setting and planning* according to the COPES-model also differed significantly between tasks: Students not only mentioned significantly fewer goals for the *remember* task than for the *evaluate* task ( $t(34) = -3.87$ ,  $p < .001$ ) but also reported qualitatively different goals ( $X^2 = 42.02$ ,  $df = 9$ ,  $p < .001$ ): For the *remember* task students most often mentioned the goals to "find the right answer" ( $n = 15$ ), to "understand the hypertext" ( $n = 11$ ) and to "find a quick solution" ( $n = 7$ ). For the *evaluate* task on the other hand, most students mentioned the goals to "write a comprehensible answer for the client" ( $n = 9$ ) or to "give an independent evaluation" ( $n = 7$ ).

A similar picture emerged for the learning strategies (*enactment* according to the COPES-model): Students not only reported using significantly fewer strategies for the *remember* task than for the *evaluate* task ( $t(34) = -2.38$ ,  $p = .023$ ), but also reported using qualitatively different strategies ( $X^2 = 40.08$ ,  $df = 8$ ,  $p < .001$ ): For the *remember* task students most often reported that they "searched for information" ( $n = 10$ ) or tried to "find the node with the answer" ( $n = 10$ ). For the *evaluate* task on the other hand, students most often reported trying to "compare nodes" ( $n = 9$ ), trying to "evaluate nodes" ( $n = 8$ ), or trying to "deeply understand nodes by reading" ( $n = 8$ ).

With regard to the *adaptation* stage according to the COPES-model, results were mixed: On the one hand, a significant number of students reported strategy revisions within tasks ( $n = 13$  for the *remember* task and  $n = 7$  for the *evaluate* task). But on the other hand the number of reported adaptations did not differ significantly ( $X^2 = 2.12$ ,  $df = 1$ ,  $p = .180$ ). Only if the direction of strategy adaptation was considered, differences could be detected: In general students downgraded their strategy for the simple *remember* task from an initial

deep processing strategy to a more superficial strategy while the reverse change could be observed for the complex *evaluate* task. Consequently, students adapted their strategies in a meaningful way to the affordances of the particular tasks. For example, for the *remember* task student AALF23 reported that she first tried to get an overview of all three DNA analysis methods presented in the hypertext and tried to understand all methods. But after a while, she gave up due to time considerations and instead used the “search” command to search for the specific answers relevant for this specific multiple-choice task. Consider another example for the *evaluate* task: Student AHAA22 reported that she first searched for the term “paternity” in the hypertext and expected to find a single hypertext page that explicitly compared the different methods of DNA analysis with regard to their adequacy for paternity testing. As soon as she realized that no such page existed, she started to navigate systematically through all chapters and tried to understand and compare the different DNA analysis methods.

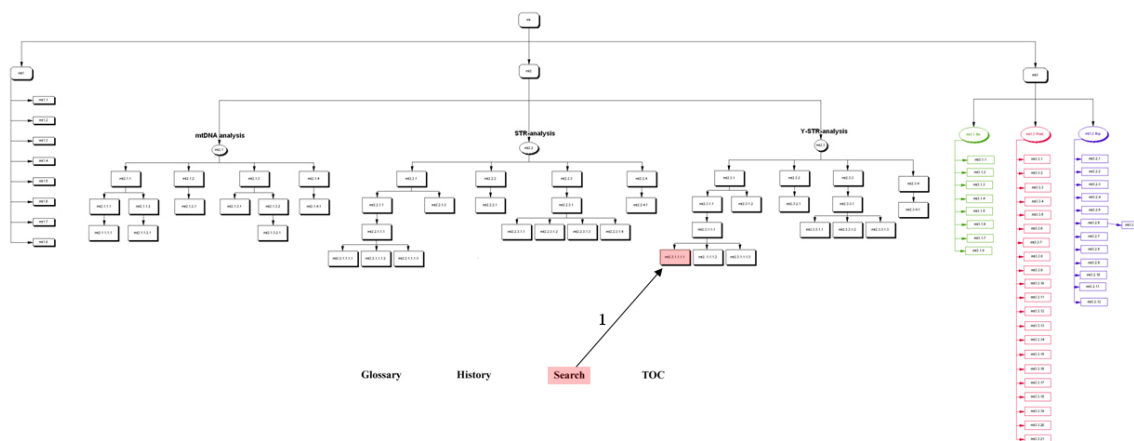


Figure 3.4-4: Navigation path of student ECJL22 for the second simple *remember* task. The accessed nodes and windows are colored red. This student went from the “search” window to the adequate target node that contained the answer and was finished with the *remember* task.

#### *Qualitative Insights from Contrastive Descriptions of Typical Approaches to Simple and Complex Tasks*

For the second simple *remember* task, most students searched for a specific term in the hypertext (e.g., for “multi-copy”), found a (almost) literal match between one sentence in the hypertext and one multiple-choice option and were finished with the task. Most “level 3” nodes were accessed for the simple *remember* task ( $M = .77$ ,  $SD = 1.09$ ; sum = 27). This fits the tasks’ requirement: The node best suited to answer this task was a “level 3” node about multi-copy Y-STR nodes. Consider an exemplary path diagram of the student ECJL22 (Figure 3.4-4) that shows that this student searched for a specific term, immediately selected the relevant target page and was finished with the second *remember* task. Students also often described this approach in their concurrent thoughts or in the retrospective stimulated recall interview (AAAD17: “... tried to find a sentence that contains the answer”).

Nonetheless, a minority of students also described that they employed (or tried to employ) a deeper approach (BJPM18: “... I have tried to understand what Y-STR loci are about”).

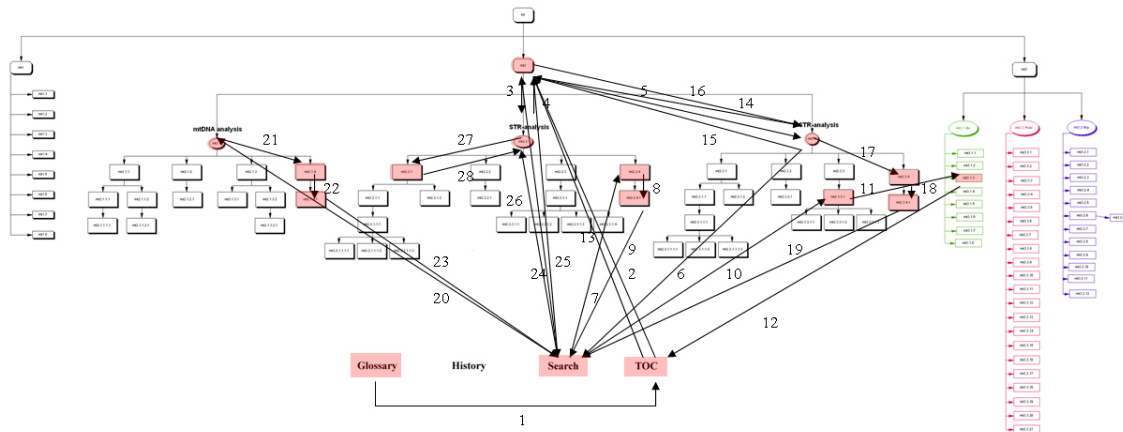


Figure 3.4-5: Navigation path of student ECJL22 for the complex *evaluate* task. The accessed nodes and windows are colored red. This student employed multiple navigational commands and accessed at least the “application” nodes for each DNA analysis method.

For the complex *evaluate* task on the other hand, students employed a variety of navigational strategies. Furthermore, most students were aware that they needed information about all different DNA analysis methods and thus had to access multiple hypertext nodes (AHAA22: “... the information is distributed across multiple chapters”). Consider an exemplary path diagram of the student ECJL22 that shows that this student accessed significantly more nodes for the *evaluate* task than for the *remember* task (compare Figure 3.4-4 and Figure 3.4-5). Furthermore, this student used multiple navigational commands (e.g., hierarchical commands such as “parent” and “child”, but also “search” and the “table of content” (“TOC”) to get to non-adjacent, more distant nodes). Thematically, this student accessed nodes from all DNA analysis methods: For each method at least the overall introductory node and all hierarchical levels of the chapter on “applications” were visited. In general, students’ concurrent thoughts and their answers in the stimulated recall interview indicate that they often processed the content very deeply (AHDD17: “... that [Y-STR analysis] is only possible if the father has a son. That is not mentioned in the text, but that has to be so, hasn’t it?”); they critically evaluated what they read (AAAD17: “I am asking myself where the gaps are, that is, when the method can not be applied ...”) and focused on critical information. This is also indicated by the fact that most “problem” nodes were accessed for this task ( $M = 1.11$ ,  $SD = 3.54$ ; sum = 39). Nonetheless, a minority of students was also satisfied with superficial understanding (BJPM18: “[the task was finished] as soon as I found the node about the reasons for the application of a STR analysis”).



To summarize the results with regard to research question one: Students' concurrent actions captured by logfiles, their concurrent thoughts captured by prompts, their number of written words, their answers to the stimulated recall interview as well as qualitative descriptions of typical task solutions of simple and complex tasks demonstrate significant *discrimination* between tasks of different complexity. Only task difficulty (percentage of correct answers) did not differ significantly between tasks of different complexity.

### 3.4.2.2 Do Students Calibrate to Task Complexity?

To answer the second research question measures of *relative calibration* were utilized: Within-subject Goodman-Kruskal Gamma correlations ( $G$ ) between the dependent variables and the *Bloom-Categories* were computed ( $n = 5$ , for the 5 tasks) and subsequently  $Z$ -transformed into *calibration* indices. These *calibration* indices were tested against zero and scrutinized with regard to their effect sizes. This method was only applied to all dependent variables that were measured for all five tasks.

Calibration indices for all dependent variables significantly differ from zero and correspond to correlations of large effect size ( $G > .50$ ). More specifically, the mean *calibration* index for the *number of accessed nodes* (NAN) is 1.45 ( $SD = 1.31$ ) which corresponds to a correlation of  $G = .90$  and significantly differs from zero ( $t(34) = 6.55, p < .001$ ). This positive association between the *Bloom-Categories* and NAN indicates that students accessed few nodes for simple tasks and an ascending number of nodes with increasing *task complexity* (Figure 3.4-3). The mean *calibration* index for the *time for task completion* (TTC) is 1.69 ( $SD = 1.24$ ) which corresponds to a correlation of  $G = .94$  and significantly differs from zero ( $t(34) = 8.04, p < .001$ ). The mean *calibration* index for the concurrent thoughts categorized as “planning” (PL) is 1.38 ( $SD = 1.70$ ) which corresponds to a correlation of  $G = .88$  and significantly differs from zero ( $t(34) = 4.83, p < .001$ ). The mean *calibration* index for the concurrent thoughts categorized as “enactment” (EN) is 2.00 ( $SD = 1.26$ ) which corresponds to a correlation of  $G = .97$  and significantly differs from zero ( $t(34) = 9.42, p < .001$ ). The mean *calibration* index for the concurrent thoughts categorized as “reflection / revision” (REV) is 1.67 ( $SD = 1.67$ ) which corresponds to a correlation of  $G = .93$  and significantly differs from zero ( $t(34) = 5.91, p < .001$ ). In all cases, the positive associations indicate that students take more *time for task completion* (TTC) and implement more “planning”, “enactment” and “revision” for complex tasks.

To summarize the results with regard to research question two: Students concurrent actions captured by logfiles as well as their concurrent thoughts captured by prompts demonstrate significant *calibration* with regard to task complexity.

### 3.4.2.3 *The Impact of Prior Domain Knowledge and Epistemological Beliefs*

Before reporting the relation between the *internal conditions* – in this case, students’ prior domain knowledge (biology students vs. humanities students) and students’ epistemological beliefs indicated by the experimental groups of the epistemological sensitization (neutral introduction vs. epistemological introduction) – and students’ *discrimination* and *calibration*, it has to be noted that the two internal conditions varied independent from each other. This was guaranteed by matching the two sub-samples that received the neutral and the epistemological introduction with regard to their prior domain knowledge and their initial epistemological beliefs.

*Are students’ prior domain knowledge and their epistemological beliefs associated with their discrimination between tasks of different complexity?*

To answer this research question, students’ prior domain knowledge (biology students vs. humanities students) and students’ epistemological beliefs (epistemological sensitization: neutral introduction vs. epistemological introduction) were included as dichotomous factors in the repeated-measure analyses across five tasks. To visualize significant results, *calibration* graphs were created.

With regard to the effects of *prior domain knowledge*, the MANOVA for students’ *concurrent actions* (NAN = *number of accessed nodes* and TTC = *time for task completion*) revealed a marginally significant multivariate main effect ( $F(2,23) = 3.23, p = .052$ ) and a marginally significant interaction with the repeated-measure factor task ( $F(8,17) = 2.43, p = .059$ ). Univariately these effects were only significant for TTC (main effect:  $F(1,24) = 6.83, p = .015$ ; interaction:  $F(4,96) = 2.52, p = .046$ ). The corresponding calibration graph (Figure 3.4-6, left) indicates that humanities students spent more time on all tasks (main effect) but one (interaction): For the last *remember* task humanities students were faster. The MANOVA for students’ *concurrent thoughts* (categorized as PL = “planning”, EN = “enactment”, and REV = “reflection / revision”) revealed a significant multivariate main effect ( $F(3,30) = 4.64, p = .009$ ) and a significant multivariate interaction with the repeated-measure factor task ( $F(12,21) = 2.61, p = .026$ ). Univariately these effects were only significant for PL (main effect:  $F(1,32) = 9.90, p = .004$ ; interaction:  $F(4,128) = 3.16, p = .016$ ). The corresponding calibration graph (Figure 3.4-6, right) indicates that humanities students engaged more frequently in “planning” processes across all tasks (main effect) and that this effect was especially pronounced for the complex *evaluate* task (interaction). The ANOVA for task difficulty (*learning outcome*) revealed a significant interaction with the repeated-measure factor task ( $F(4,23) = 3.32, p = .028$ ): The corresponding calibration graph (Figure 3.4-7, left) indicates that humanities students in general were less successful at solving tasks correctly, but solved the second *remember* task better than biology students did. To get an even more detailed insight, an additional MANOVA was calculated for all sub-scores of the scoring rubric for the complex *evaluate* task. More specifically the following dependent variables were

included: the number of mentioned *DNA analysis methods*, the number of proposed *arguments*, the score for *correct evaluations* of all DNA analysis methods, and the score for the *correct conclusions*. Additionally, the sum of all sub-scores was considered (*total score*). In order to compare these scores, all scores were standardized on a scale from 0 to 100 (100 representing 100 % of possible points). Results indicate two univariate effects: A significant effect for the *correct evaluations* ( $F(1,25) = 4.52, p = .044$ ) and a marginally significant effect for the *total score* ( $F(1,25) = 3.01, p = .095$ ). In both cases, biology students outperformed humanities students (Figure 3.4-7, right).

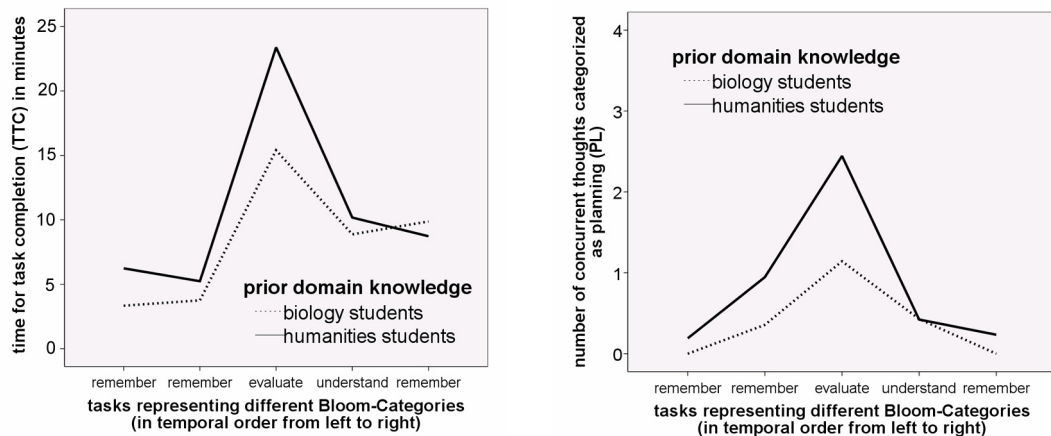


Figure 3.4-6: Calibration graphs depicting students' *time for task completion* (TTC, left) and students' "planning" (PL, right) as a function of task complexity (*Bloom-Categories*) and prior domain knowledge (biology students vs. humanities students).

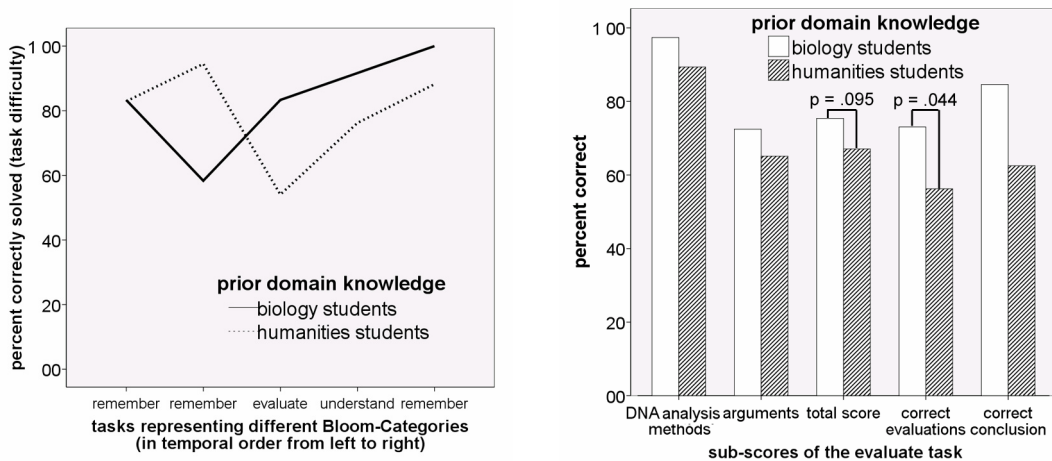


Figure 3.4-7: Calibration graph (left) depicting task difficulty as a function of task complexity (*Bloom-Categories*) and prior domain knowledge (biology students vs. humanities students); and a bar graph (right) depicting correctness on all sub-scores of the complex *evaluate* task as a function of prior domain knowledge.

With regard to the effects of the *epistemological sensitization*, the MANOVA for students' *concurrent actions* (NAN = *number of accessed nodes* and TTC = *time for task completion*) revealed a significant univariate interaction for TTC ( $F(4,96) = 3.41, p = .012$ ). The corresponding calibration graph (Figure 3.4-8, left) indicates that students who read the epistemological introduction spent less time on simple *remember* tasks, but significantly more time on the complex *evaluate* task than their counterparts who read the neutral introduction. The MANOVA for students' *concurrent thoughts* (categorized as PL = "planning", EN = "enactment", and REV = "reflection / revision") revealed a marginally significant multivariate main effect ( $F(3,30) = 2.72, p = .062$ ), but a univariately significant main effect and a univariately significant interaction for PL (main effect:  $F(1,32) = 7.62, p = .009$ ; interaction:  $F(4,128) = 3.61, p = .008$ ). The corresponding calibration graph (Figure 3.4-8, right) indicates that students who read the epistemological introduction more often engaged in "planning" processes in all tasks (main effect), and that this effect was especially pronounced for the complex *evaluate* task, still detectable for the *understand* task, but almost invisible for the *remember* tasks (interaction). *T*-test for the number of written words (*learning outcome*) revealed a marginally significant main effect ( $F(1,28) = 3.29, p = .081$ ): Students who read the epistemological introduction wrote significantly more words in both tasks with open answers (*evaluate*:  $M = 140, SD = 50$ ; *understand*:  $M = 72, SD = 29$ ) than students who read the neutral introduction (*evaluate*:  $M = 106, SD = 67$ ; *understand*:  $M = 57, SD = 24$ ). The ANOVA for task difficulty (*learning outcome*) revealed a significant main effect ( $F(1,26) = 4.53, p = .043$ ). The corresponding calibration graph (Figure 3.4-9, left) indicates that students who received the epistemological introduction performed worse across all tasks than the students who received a neutral introduction. To get an even more detailed insight, an additional MANOVA was calculated for all sub-scores of the scoring rubric for the complex *evaluate* task (see above). Results indicate a significant multivariate effect ( $F(5,21) = 4.17, p = .009$ ) that was replicated univariately for the number of mentioned arguments ( $F(1,25) = 7.19, p = .013$ ) and the correctness of the recommendation ( $F(1,25) = 7.60, p = .011$ ). While students with the epistemological introduction outperformed students with the neutral introduction on the number of mentioned *arguments*, the picture was reversed for the *correctness of the conclusion*: Students with the neutral introduction gave better recommendations (Figure 3.4-9, right).

To summarize the results with regard to research question three and *discrimination: Prior domain knowledge* has a strong impact on students' self-regulated learning process as well as on their learning outcome. While humanities students with minimal background knowledge appear to be involved in more elaborate learning processes (more TTC and more PL), biology students outperform humanities students on the learning outcome measures. *Epistemological beliefs* operationalised by the epistemological sensitization have a strong impact on students' self-regulated learning processes as well as on their learning outcome. The epistemological introduction seems to be beneficial for the learning process because students

reading this introduction regulate their time on task (TTC) more adequately and engage more frequently in “planning” (PL). Learning outcome results are mixed: While students with the epistemological introduction write more words and propose more arguments, they are less able to draw the correct conclusions.

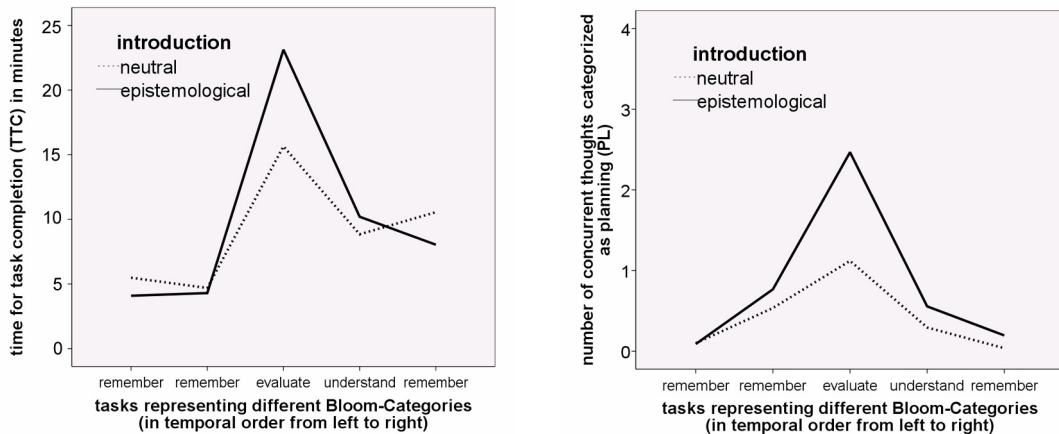


Figure 3.4-8: Calibration graphs depicting students’ *time for task completion* (TTC, left) and “planning” (PL, right) as a function of task complexity (Bloom-Categories) and the epistemological sensitization (neutral introduction vs. epistemological introduction).

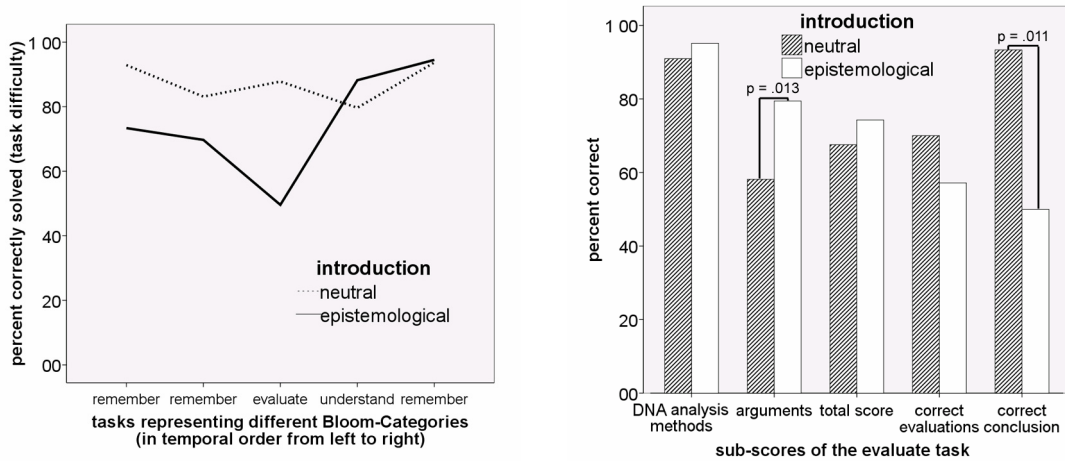


Figure 3.4-9: Calibration graph (left) depicting task difficulty as a function of task complexity (*Bloom-Categories*) and the epistemological sensitization (neutral introduction vs. epistemological introduction); and a bar graph (right) depicting correctness on all sub-scores of the complex *evaluate* task as a function of the epistemological sensitization (neutral introduction vs. epistemological introduction).

*Are students' epistemological beliefs and prior domain knowledge associated with their calibration to task complexity?*

To answer this research question, MANOVAs were computed with students' prior domain knowledge (biology students vs. humanities students) and students' epistemological beliefs (epistemological sensitization: neutral introduction vs. epistemological introduction) as factors and the *calibration indices* for all relevant variables (logfiles: NAN, TTC; concurrent thoughts: PL, EN, REV) as dependent variables.

With regard to *prior domain knowledge* only one marginally significant effect was found: Humanities students tended to calibrate their *number of accessed nodes* (NAN) better to task complexity than biology students ( $F(1,31) = 3.82, p = .060$ ; biology students:  $G = .74$ , humanities students:  $G = .95$ ). This indicates, that humanities students accessed more nodes for the *understand* task than for the *remember* tasks and more nodes for the *evaluate* task than for the *understand* task. For biology students this rank order was not as pronounced. With regard to *epistemological beliefs* only one marginally significant effect was found: Students who received the epistemological introduction tended to calibrate their *time for task completion* (TTC) stronger to task complexity than students who received the neutral introduction ( $F(1,31) = 3.02, p = .092$ ; neutral introduction:  $G = .88$ , epistemological introduction:  $G = .97$ ; see Figure 3.4-8, left). This indicates that the epistemological introduction led to a rank ordering of tasks with regard to processing time analogous to the *Bloom-Categories* while this was not as strongly the case for the neutral introduction.

To summarize these results with regard to research question three and *calibration*: Prior domain knowledge was detrimental for calibration while the epistemological introduction was beneficial for calibration.

#### **3.4.2.4 Determinants of the Learning Outcome**

The learning outcome can be defined by the number of correctly solved tasks (overall learning outcome score), by the success at one specific task (the second *remember* or the *evaluate* task), or by the sub-scores of the scoring rubrics (*evaluate* task). The impact of different quantitative variables (internal conditions, see previous section; students' calibration; and additional variables capturing their concurrent learning process) on these indicators of learning outcome will be scrutinized statistically. Additionally, a qualitative answer will be attempted by analyzing the learning process of the most and least successful students.

##### *Determinants of the Overall Learning Outcome Score*

An overall learning outcome score was computed by summing up the number of correctly solved tasks (ranging from 0 to 5 points). On average, students scored 4.03 points ( $SD = .82$ ) which corresponds to 81 %. A number of potential predictor variables were correlated with this score: the overall *number of accessed nodes* (NAN), the overall *time for task*

*completion* (TTC), the overall number of concurrent thoughts categorized as “planning” (PL), “enactment” (EN) and “reflection / revision” (REV); the *calibration* indices for NAN, TTC, PL, EN, and REV; the duration of and number of node accesses for all constituent parts of the hypertext (“level 1”, “level 2”, “level 3”, “biological background”, “examples”, and “problems”). Results indicate only one significant correlation: The *calibration* indices for the *time for task completion* (TTC) were negatively correlated with the overall learning outcome score ( $r = -.52, p = .004, n = 29$ ). Thus, students who better adapted their *time for task completion* to the complexity of the learning tasks solved fewer tasks correctly.

#### *Determinants of Success on the Second Remember Task*

The second simple *remember* task was solved correctly by 77 % of all students ( $n = 27$ ). A number of potential task specific predictor variables were compared between students who solved this task correctly and those who did not solve it correctly: *number of accessed nodes* for task number two (NAN), *time for task completion* for task number two (TTC), number of concurrent thoughts categorized as “planning” (PL), “enactment” (EN), and “reflection / revision” (REV) for task number two; the duration of and number of node accesses for all constituent parts of the hypertext for task number two. Results indicate only one significant effect: Students solving this task correctly significantly spent more time on “level 3” ( $M = 1:39, SD = 1:50, n = 27$ ) than students who were not able to solve this task correctly ( $M = 0:16, SD = 0:39, n = 8; t(33) = -2.07, p = .046$ ).

#### *Determinants of Success for the Evaluate Task*

For the complex *evaluate* task, the correctness of the conclusion and the total score (including the quality of argumentation) were considered as indicators of learning outcome. For both outcome measures a number of potential task specific predictor variables were considered: *number of accessed nodes* for the *evaluate* task (NAN), *time for task completion* for the *evaluate* task (TTC), the number of concurrent thoughts categorized as “planning” (PL), “enactment” (EN), and “reflection / revision” (REV) for the *evaluate* task; the duration of and number of node accesses for all constituent parts of the hypertext for the *evaluate* task.

With regard to the *correctness of the conclusion*, 60 % of all students ( $n = 21$ ) recommended STR analysis for paternity testing, thus giving the correct recommendation. The predictor variables were compared between students who solved this task correctly and those who did not solve it correctly. Three effects were detected: Students who gave the correct recommendation significantly processed these task shorter (TTC:  $t(33) = 2.45, p = .020$ ; correct:  $M = 20:11, SD = 10:07, n = 21$ , incorrect:  $M = 29:23, SD = 11.55, n = 14$ ), they uttered fewer concurrent thoughts categorized as “planning” (PL:  $t(33) = 2.48, p = .019$ ; correct:  $M = 1.38, SD = 1.32, n = 21$ , incorrect:  $M = 2.78, SD = 2.04, n = 14$ ), and they processed “level 1” shorter ( $t(33) = 2.13, p = .041$ ; correct:  $M = 8:08, SD = 4:40, n = 21$ , incorrect:  $M = 13:00, SD = 8.48, n = 14$ ).

With regard to the *total score for the evaluate task* that also captures the quality of argumentation, students on average received 10.06 ( $SD = 2.47$ ) out of 15 points for their written answers (67 % of points). The predictor variables were correlated with this score. Six effects were detected: Longer processing of this task (TTC:  $r = .41, p = .015, n = 35$ ) and more concurrent thoughts categorized as “enactment” (EN:  $r = .30, p = .080, n = 35$ ) were positively associated with a higher score. More detailed data from the logfiles revealed that especially the number of ( $r = .31, p = .068$ ) and duration of ( $r = .29, p = .89$ ) accessed “problem” nodes as well as the number of ( $r = .36, p = .035$ ) and duration of ( $r = .33, p = .056$ ) accessed “example” nodes were positively related to the total score of the *evaluate* task.

#### *Qualitative Analysis of the Most and the Least Successful Student*

Success can be defined with regard to the overall learning outcome score (see above) or with regard to the total score for the *evaluate* task (see above). The patterns of these two indicators of success partly differ, especially for the least successful students. Consider the example of student MGJH20. This student started out with low prior knowledge (4 points on the molecular biology knowledge test) and only managed to give the correct answers for two out of five tasks (40 %), but was quite successful at writing an adequate answer for the complex *evaluate* task (11 points, 73 %). For most students however, both indicators of success are fairly congruent. Two typical examples, one of a successful student and one of a less successful student will be described contrastively in further detail for the second *remember* task and for the *evaluate* task.

Student GMMT08 is a humanities student with moderate background knowledge in molecular biology (4 points in the prior knowledge test) who read the neutral introduction. With regard to the overall learning outcome, she answered 3 out of 5 tasks correctly (60 %) and received 4 points for her written answer to the *evaluate* task (27 %). She solved the second *remember* task correctly, but did not give the correct recommendation for the *evaluate* task. Only one student performed worse with regard of correctness, and GMMT08 performed worst in the whole sample with regard to the written answer to the *evaluate* task.

As a more successful example, consider student HSH05. She also is a humanities student with even less background knowledge in molecular biology (2 points in the prior knowledge test) who read the epistemological introduction. With regard to the overall learning outcome, she answered all 5 tasks correctly (100 %) and received 13 points for her written answer to the *evaluate* task (87 %). She solved the second *remember* task correctly and gave the correct recommendation for the *evaluate* task. She was one out of nine students who answered all tasks correctly, furthermore only three students gained more points for their written answer to the *evaluate* task.

What are the relevant similarities and differences between these two students: Why did they perform so differently? In order to answer these questions, consider these students' navigation paths (Figure 3.4-10 and Figure 3.4-11). Both students used a similar number of



navigational moves: For the *remember* task GMMT08 used 4 moves and HSH05 used 3 moves and for the *evaluate* task GMMT08 used 26 moves and HSH05 used 27 moves. Furthermore, both students accessed nodes from similar hypertext chapters: For the *remember* task both students accessed nodes from the chapter on Y-STR analysis and for the *evaluate* task both students accessed nodes from all chapters on mtDNA analysis, STR analysis, and Y-STR analysis. Additionally, both students frequently re-accessed nodes for the complex *evaluate* task. Another similarity concerns the use of navigational commands: Both students demonstrated flexibility in their navigation and not only used hierarchical commands such as “parent” or “child” but also advanced features to get to more distant nodes (e.g., the “TOC” (table of content) or the “search” function).

As both students successfully solved the *remember* task, only differences with regard to the *evaluate* task will be discussed in detail. On the *evaluate* task students differed on the following variables: The least successful student GMMT08 accessed 12 nodes in total whereas the most successful student HSH05 accessed only 8 nodes. Thus, the more successful student re-accessed these nodes more often. Furthermore, these nodes are differently distributed across different hypertext parts indicating that students focused on different content. While the least successful student GMMT08 accessed 4 nodes from the introduction, 1 node from the chapter on mtDNA analysis, 2 nodes from the chapter on STR analysis, 4 nodes from the chapter on Y-STR analysis, the most successful student HSH05 accessed 1 node from the introduction, 2 nodes each from the chapter on mtDNA analysis and STR analysis and 3 nodes from the chapter on Y-STR analysis. Thus, the least successful students clearly almost exclusively focused on Y-STR analysis and the – almost totally irrelevant – introduction, the most successful student clearly focused on all DNA analysis methods equally. Another difference in students’ focus concerns not only the distribution across different hypertext parts, but also different sub-chapters and levels within chapters. The least successful student GMMT08 accessed different nodes in different chapters: For example, in the chapter on mtDNA analysis, only the overall introductory node was accessed, in the chapter on STR analysis both nodes on “applications” were accessed and in the chapter on Y-STR analysis nodes from the “basic idea”, the execution “in the lab”, and the “interpretation” were accessed. The most successful student HSH05 on the other hand accessed the same nodes in each chapter: Both “application” nodes were accessed in the chapters on mtDNA analysis, STR analysis, and Y-STR analysis. In the chapter on Y-STR analysis the overall introductory node was accessed additionally. Thus, the successful student would have the opportunity to explicitly compare all DNA analysis methods on one dimension (i.e., “applications”), while the least successful student did not create such an opportunity. The navigation path of the most successful student HSH05 also indicates that she probably systematically compared the DNA analysis methods with regard to their “applications”: She accessed and re-accessed the “application” nodes from all DNA analy-

sis methods repeatedly in the following order: Y-STR analysis, STR analysis, mtDNA analysis, STR analysis, Y-STR analysis, STR analysis, Y-STR analysis, and mtDNA analysis.

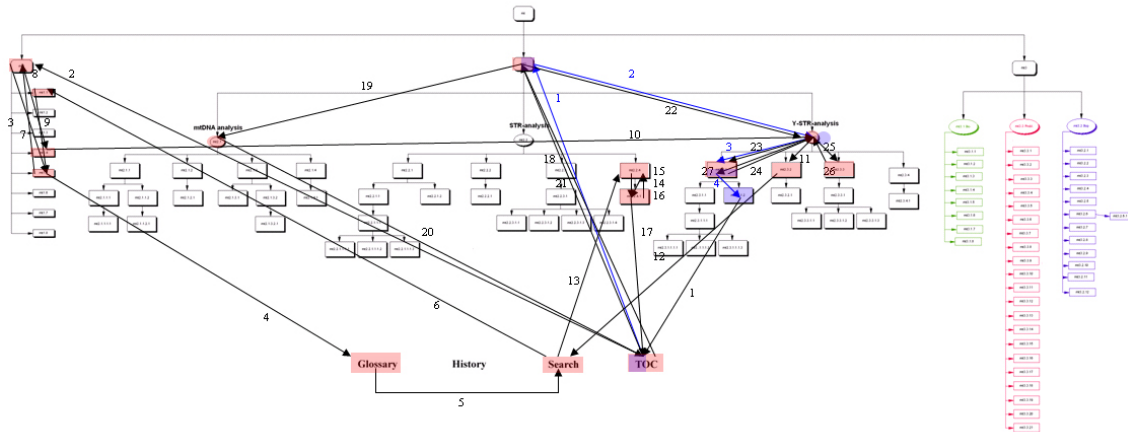


Figure 3.4-10: Navigation paths of student GMMT08. All navigational moves for the *remember* tasks are marked blue with blue colors indicating visited nodes. All navigational moves for the *evaluate* tasks are marked black with pale red colors indicating visited nodes.

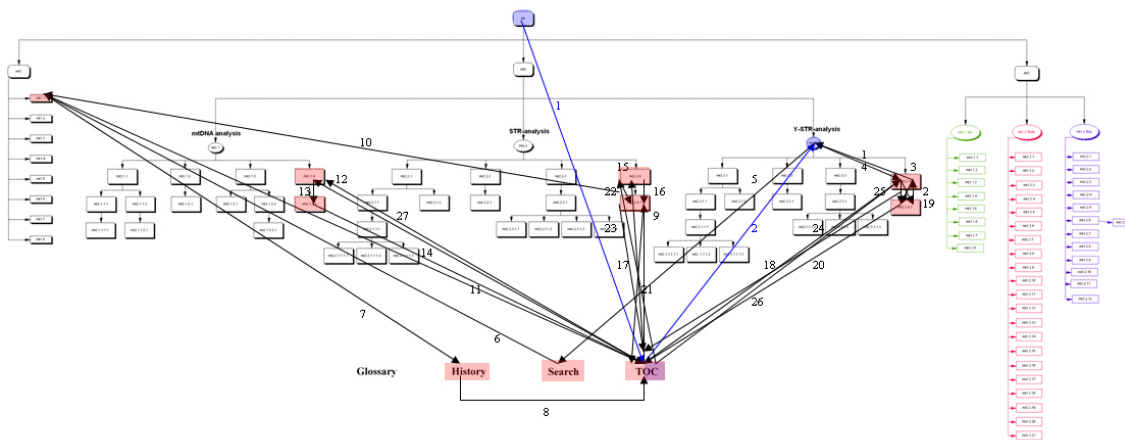


Figure 3.4-11: Navigation paths of student HHSH05. All navigational moves for the *remember* tasks are marked blue with blue colors indicating visited nodes. All navigational moves for the *evaluate* tasks are marked black with pale red colors indicating visited nodes.

### Short Summary of Results

The results of the quantitative analyses indicate that calibration (effects of TTC on the overall learning outcome score), “planning” and spending time on top-level nodes are detrimental (effect for the correctness of the *evaluate* task) but spending time on “level 3” nodes is helpful (effect for the correctness of the *evaluate* task) if the *correctness* of solution is

considered. The *quality of the written answers* on the other hand is especially related to spending more time on the *evaluate* task, to “enacting” more strategies, and to more and longer accesses of “problem” and “example” nodes that elaborate the content of the main hypertext. Additional qualitative results indicate that successful students are better at differentiating between relevant and irrelevant hypertext nodes and more systematically compare content from related nodes.

### 3.4.3 Local Discussion

The main goal of this study was to extend the results of study I focusing on the *preparatory* stage to the *enactment* stages of learning. Students solved five learning tasks of different *Bloom-Categories* with the hypertext on “genetic fingerprinting”. Students’ concurrent actions, their concurrent thoughts, their answers to the tasks, and their retrospective perceptions of their own learning process were captured as dependent variables. Thus, this study thoroughly investigated the *enactment* stages of the COPES-model.

#### 3.4.3.1 Discrimination and Calibration

Students consistently and significantly demonstrated their ability to adapt their judgments to the complexity of the learning tasks of different *Bloom-Categories*: The analyses concerning students’ *discrimination* and *calibration* consistently indicate positive results.

More specifically, analyses with regard to students’ *discrimination* indicate significant differences between tasks of different *Bloom-Categories* for all dependent variables capturing students’ concurrent actions and concurrent thoughts. Additionally, the qualitative data of the retrospective recall interview and qualitative analyses of students’ learning processes point in the same direction. The results with regard to the *leaning outcome* (students written answers) on the other only partly support this picture: While students wrote significantly more words for complex tasks than for the simpler tasks, no significant difference could be detected with regard to task difficulty (percentage of correct answers). This lack of effect might be interpreted as a further indicator of the success of students’ adaptation task complexity: The superficial strategies enacted for the simple tasks led to successful task solution and the deep elaboration strategies enacted for the complex task also led to successful task solution. With regard to students’ *calibration* to task complexity (*Bloom-Categories*) students demonstrated good *relative calibration*: All calibration indices (concurrent actions and thoughts) differ significantly from zero and correspond to correlations of large effect size. All analyses consistently indicate that students spend more time on complex tasks, access a higher number of nodes for complex tasks and show an increase in all types of concurrent

thoughts (“planning”, “enactment”, and “reflection / revision”). These findings may seem trivial at first glance because more time spent on a task automatically leads to a higher number of prompts per task. Nonetheless students’ could have only increased their (superficial) “enactment” without increasing their metacognitive involvement (“planning” and “reflection / revision”; also see Stahl, Bromme, Stadler, & Jaron, 2006). However, these results show that students increased their concurrent *cognitive* as well as *metacognitive* processing for more complex tasks. Thus, students metacognitively monitored task complexity and adapted their learning process accordingly.

These results are consistent with the COPES-model (Winne & Hadwin, 1998; chapter 2.1) that assumes that students systematically adapt their learning process to external conditions. Additionally, these results are mostly consistent with those of previous empirical studies about task complexity (chapter 2.3; Gall, 2006; Klayman, 1985; Rouet, 2003; Winne & Jamieson-Noel, 2003). These studies consistently indicate that learners demonstrate good self-regulation for simple tasks but less adequate self-regulation for complex tasks. The results of this study are consistent because in both cases learners process differently complex tasks differently and systematically adapt their behavior to task complexity. However, the results of this study do not say anything about the absolute quality of students’ self-regulation for simple or complex tasks.

#### ***3.4.3.2 The Impact of Prior Domain Knowledge and Epistemological Beliefs***

With regard to the impact of *internal conditions* on students’ *discrimination* and *calibration*, students’ prior domain knowledge (biology students vs. humanities students) as well as their epistemological beliefs (epistemological sensitization: neutral introduction vs. epistemological introduction) demonstrated significant impact.

##### *Prior Domain Knowledge*

With regard to prior domain knowledge, the effects are mostly consistent: Students’ concurrent actions indicate that biology students were faster across all tasks (TTC) and calibrated their *number of accessed nodes* (NAN) not very strongly to task complexity. Their concurrent thoughts indicate that biology students also employed less “planning” for all tasks (PL). Still, biology students outperformed humanities students with regard to the learning outcome. Most likely, biology students knew more technical terms and concepts needed to thoroughly understand the content of the hypertext. Therefore, they might have been able to faster comprehend the content (TTC). Prior knowledge might also have enabled them to access fewer nodes compared with humanities students who probably accessed additional nodes in order to gather the necessary knowledge to answer the tasks. Because especially the complex *evaluate* task afforded reading multiple nodes, this might have resulted in a higher calibration of *number of accessed nodes* (NAN) for humanities students. Additionally,

biology students' metacognitive strategies might have been automated while humanities students might have tried to compensate their lack of knowledge by additional metacognitive strategies thus explaining their enhanced "planning" (PL). However, despite humanities students' apparently deeper processing their strategy was not enough to compensate for biology students' initial advantage (learning outcome).

However, some small inconsistencies with this overall pattern were detected that need special attention: First, the *calibration* graph for the effect of prior domain knowledge on *time for task completion* TTC (Figure 3.4-6) indicates that biology students were slower on the last *remember* task while they were faster on all other tasks. This effect might be an artifact of the experimental procedure: For all students the maximal time for learning was fixed. Due to their longer processing of the first four tasks, humanities students most likely ran out of time for the last task resulting in shorter TTC. Biology students might have been finished before they ran out of time but forgot to log out resulting in longer TTC. Second, biology students demonstrated superior learning outcome for all tasks but for the second *remember* task where humanities students excelled (Figure 3.4-7). This effect might be explained by referring to the tasks' content. It concerns multiple-copy Y-STR loci and the correct answer involves recognition of the fact that two alleles could be detected at these loci. However, normal Y-chromosome loci only have one allele per locus. Multiple-copy loci are a special case in which two alleles (instead of the normal one) are visible in the electrophoretogram. Because of their prior domain knowledge, biology students might have immediately realized that all Y-chromosome loci could only have one allele and thus could have rejected the correct option. Humanities students on the other hand read in the introduction that all (autosomal) loci possess two alleles. Thus, they might have immediately selected the only option involving two alleles, albeit for the wrong reasons.

The detected effects are consistent with the predictions of the COPES-model (Winne & Hadwin, 1998; chapter 2.1): Prior domain knowledge impacted students' operations (captured by their concurrent actions), their metacognitive processes (captured by their "planning") and their product of learning (captured by their learning outcome). Additionally, these results are mostly consistent with previous empirical studies about the impact of prior domain knowledge (chapter 2.5). For example, in accordance with the presented results all of these other studies demonstrate the benefit of prior domain knowledge for the leaning outcome (Ford & Chen, 2000; Lawless, Brown, Mills, & Mayall, 2003; Lind & Sandmann, 2003; Rouet, 2003; Rouet, Favart, Britt, & Perfetti, 1997). Additionally, some empirical studies found that knowledge was especially beneficial when dealing with complexity (Calisir & Gurel, 2003; McDonald & Stevenson, 1998; McNamara, Kintsch, Songer, & Kintsch, 1996; Möller & Müller-Kalthoff, 2000; Potelle & Rouet, 2003; Salmerón, Kintsch, & Canas, 2006) which is consistent with the interaction effects of this study demonstrating the most pronounced impact of prior domain knowledge for the complex *evaluate* task. And one empirical results obtained with the methodology of the traditional calibration paradigm indi-

cates that more prior domain knowledge might result in less calibration which is also consistent with the presented results (Glenberg & Epstein, 1987a). However, most of the other empirical studies indicate the benefit of prior domain knowledge for the use of more elaborate learning strategies (Ford & Chen, 2000; Lawless, Brown, Mills, & Mayall, 2003; Lind & Sandmann, 2003) and more metacognitions (Priest & Lindsay, 1992; Veenman & Elshout, 1999) which is clearly inconsistent with results of this study.

#### *Epistemological Beliefs Operationalised by the Epistemological Sensitization*

With regard to the epistemological sensitization the effects are mostly consistent: Students' concurrent actions indicate that students who read the epistemological introduction spent more time on the *evaluate* task thus demonstrating better calibration on this dependent variable (TTC). Their concurrent thoughts indicate that students who read the epistemological introduction also employed more "planning", especially for the *evaluate* task (PL). With regard to the learning outcome, reading the epistemological introduction was associated with more written words and superior argumentation, but was detrimental for overall success in terms of correctness. Most likely, the epistemological introduction triggered an awareness of the complexity, uncertainty, and variability of knowledge. This awareness might have elicited standards of deeper learning and thus a more elaborate self-regulated learning process including better adaptation to task complexity. The additional resources spent on the complex *evaluate* task were probably utilized for processes of critical evaluation or the attempt to grasp the interrelated nature of the presented information. Additionally, students might have tried to convey their new insights by writing adequately complex answers. However, this increased awareness might also have made it more difficult for students to finally decide on one single correct answer: Some comments in the epistemological introduction stressed the multiplicity of possible opinions and the difficulty to find final truths. Thus, students might have experienced confusion and might have been easily misled to accept wrong answers. The students who read the neutral introduction on the other hand were mainly presented with factual background information. Most likely, they did not ponder on the nature of the underlying knowledge, just accepted it at face value.

These effects are consistent with the predictions of the COPES-model (Winne & Hadwin, 1998; chapter 2.1): The epistemological sensitization that elicited more "sophisticated" beliefs impacted students' operations (captured by their concurrent actions), their metacognitive processes (captured by their "planning") and their product of learning (captured by their learning outcome). Additionally, these results are mostly consistent with those of previous empirical studies (chapter 2.6). Most empirical studies support the notion of a beneficial main effect of "sophisticated" epistemological beliefs on learning strategies consistent with the main effect for "planning" in this study (Bartholomé, Stahl, Pieschl, & Bromme, 2006; Bendixen & Hartley, 2003; Jacobson & Spiro, 1995; Kardash & Howell, 2000). The interaction effects detected for *time for task completion* and "planning" in this

study were not yet detected in any previous empirical study, but were predicted theoretically based on a flexible notion of “sophisticated” epistemologies (Hammer & Elby, 2002). Additionally, the fact that students with the epistemological introduction wrote better argumentations is also consistent with previous empirical results (Kardash & Scholes, 1996; Mason & Boscolo, 2004; Mason & Scirica, 2006; Schommer, 1990). However, the poor performance of the students with the epistemological introduction with regard to correctness seems inconsistent with empirical results demonstrating the general benefit of “sophisticated” beliefs for the learning outcome (Schommer, 1990; Schommer, Crouse, & Rhodes, 1992). Note though, that most of these results were obtained with complex learning outcome measures and that the results are more mixed for simpler learning outcome measures comparable to the correctness of answers: For example, Mason and Boscolo (2004) detected no impact of epistemological beliefs on students’ text comprehension.

### ***3.4.3.3 Determinants of the Learning Outcome***

As elaborated in the previous section, more prior domain knowledge was consistently beneficial for the learning outcome while the epistemological introduction was beneficial for qualitative indicators of learning outcome but detrimental for correctness.

Additionally, calibration with regard to *time for task completion* (TTC) proved detrimental for students’ learning outcome indicated by the overall correctness of their answers. Consistently, students who spend less time on the evaluate task were more likely to give the correct recommendation for the complex *evaluate* task. Furthermore, spending much time on “level 1” and giving many “planning” (PL) answers to prompts were also negatively related to proposing correct conclusions for the *evaluate* task. Thus, all of these results consistently seem to indicate that deep processing is detrimental for performance, especially for more complex tasks. These counterintuitive results might be explained by considering that these results are highly task specific and highly dependent on the measurement of learning outcome. Depending on the task (*task specificity*), more TTC could either be beneficial or detrimental: Opposed to the above-mentioned results, more time on “level 3” was associated with giving correct answers for the second *remember* task, probably because the target information about multi-copy Y-STR loci could be found on a “level 3” node. And depending on the measurement of success (*measurement of learning outcome*), contradictory predictors of success could be determined: Opposed to the above-mentioned results, more time on “examples” and “problems” was beneficial for the total score of the *evaluate* task also considering more qualitative sub-scores. If both aspects of success are considered equally (*correctness* and *quality*), deep processing strategies seem to be beneficial for success for the complex *evaluate* task as indicated by a qualitative comparison between the most and the least successful students. More successful students more systematically accessed and re-accessed a small number of relevant hypertext nodes to systematically compare all DNA

analysis methods, probably to engage in processes of structure mapping (Gentner & Markman, 1994). Less successful students on the other hand accessed more irrelevant hypertext nodes and were less systematic in their navigation behavior.

These effects are consistent with the predictions of the COPES-model (Winne & Hadwin, 1998; chapter 2.1) insofar as the learning processes (operations like TTC) were related to the product of learning (learning outcome) and insofar as more elaborate operations mostly elicited better learning outcome. However, the COPES-model also predicted that pronounced calibration should result in better learning outcome and the only detected effect pointed in the reverse direction.

#### ***3.4.3.4 Open Issues and Implications***

This study was successful at demonstrating that students in fact do *discriminate* between tasks of different complexity and *calibrate to task complexity* and that these processes are impacted by students' *prior domain knowledge* and *their epistemological beliefs* in the *enactment* stages of learning (previous sections). However, some limitations have to be considered.

##### *Open Issues and Limitations*

In this section most involved variables will be scrutinized with regard to their validity, corresponding limitations, and open issues. Note though that some variables were already discussed in the previous sections (e.g., the different measures of learning outcome in the section about the “determinants of the learning outcome”).

First of all, the small sample size limits the generalizability of the results. An additional argument in this direction concerns the rather qualitative and explorative focus of this study. For example, students' *concurrent thoughts* and their answers in the *retrospective stimulated recall interviews* might have been interpreted subjectively. To counteract this potential problem, expert discussions and multiple raters were used to enhance objectivity of the coding schemes and the coding procedures. Empirical results indicate that this might have been quite successful as concurrent validity seems to be given: The results from these more qualitative sources are congruent with the results from more quantitative sources (logfiles). Additionally, consider an advantage of this qualitative focus: These data sources give rich insight into students' whole self-regulated learning process.

A second issue concerns the measures extracted from the logfiles. Concerning logfile-analyses, single-unit measures (Richter, Naumann, & Noller, 2003) such *time for task completion* (TTC) or *number of accessed nodes* (NAN) are relatively easy to capture. Nonetheless, these single-unit measures only give a rough first impression but do not reflect the wealth of information contained in the original logfiles. Furthermore, they do not tell anything about the navigational commands used by students and their sequence of visited nodes. Path diagrams on the other hand visualize this detailed information (Figure 3.4-11), but the patterns



are hard to compare systematically (Barab, Bowdish, Young, & Owen, 1996; Canter, Rivers, & Storrs, 1985; McEneaney, 2001). Note though that all path diagrams of this study indicate that students were deeply involved in task solutions and did not enact “feature seeking” or “apathic” strategies (Lawless, Brown, Mills, & Mayall, 2003; Lawless & Kullikowich, 1996). Note that the measures extracted from the logfiles (TTC and NAN) seem appropriate for two reasons: First, so far there are only few studies investigating, for example, the impact of epistemological beliefs on *online* learning processes. And second, these measures seem to be at an adequate *level of granularity* to detect effects.

A third open issue concerns the *absolute* quality of students’ calibration. Even though students in general are quite successful at discriminating between and calibrating to tasks of different complexity they might be still far from perfect. As highlighted in the discussion of study I it would not have dramatic effects to overestimate the complexity of a simple task, but underestimating the complexity of a complex task could have detrimental effects. Data from this study tentatively indicates that students might in fact underestimate the complexity of complex tasks. First, even though not significantly so, simpler tasks were more often solved correctly (77 % – 93 %) than the complex *evaluate* task (60.0 %). Second, logfiles show that most students did not access all relevant nodes for the complex *evaluate* task, thus their understanding of the presented DNA analysis methods can only be fragmentary and incomplete. Third, students also explicitly reported that they noticed this problem, but had to give superficial answers in order to adhere to the time constraint: For example, student AHDD17 stated “I have given it [the answer] very shortly [...] but you could write much more about this topic [...], ten pages”.

### *Implications*

Because of the above-mentioned open issues only tentative implications can be drawn. One tentative implication concerns adequate scaffolds for students’ whole self-regulated learning process – not only for the preparatory stages of learning (study I). For example, the epistemological sensitization implemented in this study elicited deeper learning processes and better written argumentations. Thus, such an epistemological sensitization that focuses students on the epistemology of a domain seems a good way to foster more adequate learning processes. However, other scaffolds are feasible. If students in fact try to adapt to a multitude of *external conditions* as assumed in the COPES-model, it could be helpful to focus them on the most relevant external conditions (e.g., on the complexity of the task at hand). Thus, instructing students adequately about task demands or about scoring criteria applied to their written answers could potentially also enhance adequate self-regulation.

Another potential implication concerns the long-term effects of the epistemological sensitization implemented to foster more adequate self-regulated learning. This intervention led to fewer correctly solved questions in this study, but it also led to better argumentation. The latter can be considered a more desirable goal of education than the former. Still, this

differential effect is interesting. It indicates that short-term interventions such as this epistemological sensitization might not be recognized as beneficial if inadequate criteria are considered. Additionally, the detrimental effect on the correctness might be an initial effect which might reverse if a long-term intervention was implemented. In general, students who pass through a phase of epistemological change experience some confusion (chapter 2.6.1). Thus, they might have initial difficulty with regard to correct answers. These problems might disappear, however, as soon as students' confusion disappears and they reach a more "sophisticated" stage of epistemological thinking. Thus, it would be an interesting question to investigate the long-term effects of an instructional intervention strongly focusing on the epistemology of the educational domain.

### 3.5 Study III<sup>7</sup> – Calibration to Text Complexity in the Enactment Stages

This study pursued the following research questions: (1) Do students *discriminate* between texts of different complexity (different hierarchical hypertext levels)? (2) Do students *calibrate* to *text complexity*? (3) Do the *internal conditions* prior domain knowledge and epistemological beliefs impact these processes of *discrimination* and *calibration*? (4) What factors determine the learning outcome (e.g., students' *calibrations* or *internal conditions*)?

Biology students ( $n = 25$ ) and humanities students ( $n = 26$ ) who varied in their epistemological beliefs learned with a hierarchical hypertext about the topic "genetic fingerprinting". During their learning processes, logfiles and questionnaire data were collected. Results indicate that students *discriminate* between texts of different complexity and *calibrate* their learning process to text complexity (e.g., by processing more complex deeper-level nodes longer). Furthermore, these discrimination and calibration processes were significantly related to their prior domain knowledge and their epistemological beliefs (*internal conditions*). Prior domain knowledge mostly impacted students' evaluation of the texts: Biology students considered all nodes more comprehensible than humanities students. With regard to epistemological beliefs, more "sophisticated" beliefs were associated with a selective reading strategy: Students processed single nodes shorter, but more nodes in total than "naïve" students. These effects were especially pronounced on deeper hierarchical hypertext levels. Additionally, more prior domain knowledge and more "sophisticated" epistemological beliefs were also beneficial for the learning outcome.

---

<sup>7</sup> Parts of the presented data were already published in Pieschl, Stahl, & Bromme (2006) and Pieschl, Stahl, & Bromme (2008, the original publication is available at [www.springerlink.com](http://www.springerlink.com)). To account for the overall research questions of this thesis, additional analyses were run and the results will be presented in more detail.

### 3.5.1 Method

#### 3.5.1.1 Procedure

All students received 15 € reimbursement for participation. To investigate the impact of prior domain knowledge, students were selectively recruited to ensure two levels of knowledge: Biology students ( $n = 25$ ) were recruited during regular courses in biology, humanities students ( $n = 26$ ) were recruited by a posting at the psychological institute.

This study was conducted in small group sessions with a maximum of 6 students per group which lasted about two hours (all materials of this study can be found in the appendices C1 – C11). First, each student completed questionnaires on demographics, their domain-general (WKI) and domain-dependent (CAEB) epistemological beliefs and their prior knowledge in molecular biology. Then all students were introduced to the structure and navigational options of the hypertext chapter about mtDNA analysis. For the main part of this study, students were instructed to read an eight-node introduction and subsequently to “learn as much as possible” about the topic of mtDNA analysis. Explicitly such an unspecific learning task was chosen to investigate students’ task interpretation as indicated by their spontaneous navigation. During this approximately one hour long learning phase detailed logfiles were automatically collected to capture students’ concurrent navigation in the hypertext. Additionally, students also had to give comprehensibility ratings for each hypertext node they read. After this learning phase students completed a knowledge test on mtDNA analysis.

In all analyses regarding the impact of *internal conditions* prior domain knowledge was included as a factor (biology students vs. humanities students) and students’ epistemological beliefs captured by the questionnaires WKI and CAEB were included as covariates.

#### 3.5.1.2 Participants

Although the advanced students of biology ( $n = 25$ ) were no experts in the specific topic of genetic fingerprinting they can be considered discipline experts (chapter 2.5.3). These biology students (13 males, 12 females) were on average 22 years old ( $SD = 1.53$ ) and studied in the 4.5<sup>th</sup> semester biology or related majors ( $SD = 1.05$ ). They already attended 9 of 10 relevant courses in molecular biology. Adequate background knowledge in molecular biology was verified by the results of a short knowledge test ( $M = 7.68$ ,  $SD = .69$ , with 8 points maximum; the knowledge test can be found in appendix C4). Their interest ( $M = 4.28$ ,  $SD = .84$ , on a 5-point scale with 1 = very low and 5 = very high) and self-rated prior domain knowledge in molecular biology ( $M = 3.24$ ,  $SD = .66$ , on a 5-point scale with 1 = very low and 5 = very high) were also quite high.

Students of humanities ( $n = 26$ ) can be considered novices (chapter 2.5.3). These humanities students (10 male, 16 females) were on average 24 years old ( $SD = 2.97$ ) and studied in the 6<sup>th</sup> semester ( $SD = 3.92$ ) a humanity major such as psychology, history or sociology. They did not attend any of the 10 relevant university courses in molecular biology. Low background knowledge in molecular biology was verified by the results of a short knowledge test ( $M = 2.77$ ,  $SD = 1.73$ , with 8 points maximum). Their interest ( $M = 2.54$ ,  $SD = .95$ , on a 5-point scale with 1 = very low and 5 = very high) and self-rated prior domain knowledge in molecular biology were also quite low ( $M = 1.81$ ,  $SD = .90$ , on a 5-point scale with 1 = very low and 5 = very high).

The difference between biology students and humanities students with regard to their prior domain knowledge was significant on all relevant variables (points in the molecular biology test:  $t(49) = -13.23$ ,  $p < .001$ ; number of attended courses relevant to molecular biology:  $t(49) = -16.62$ ,  $p < .001$ ; self-rated prior domain knowledge:  $t(49) = -6.47$ ,  $p < .001$ ). In all instances, biology students displayed more prior domain knowledge. Thus, prior domain knowledge indicated by these quasi-experimental groups (biology students vs. humanities students) will be used as dichotomous independent variable.

### 3.5.1.3 Materials

#### *Epistemological Beliefs Questionnaires*

A combination of two instruments was used to measure epistemological beliefs: The WKI (Wood and Kardash Instrument; Wood & Kardash, 2002; appendix C2) captures students' domain-general beliefs about the nature of knowledge and knowing by denotative statements. All items that do not refer to epistemology in a strict sense were eliminated (i.e., items from the dimensions *characteristics of successful students* and *speed of knowledge acquisition*). The remaining 25 items were distributed as follows: 11 items belong to the original factor *structure of knowledge* (sample item: "I like information to be presented in a straightforward fashion; I don't like having to read between the lines."), 11 items belong to the original factor of *knowledge construction and modification* (sample item: "Today's facts may be tomorrow's fiction."), and 3 items belong to the original factor *attainability of truth* (sample item: "Scientists can ultimately get to the truth."). The CAEB (Connotative Aspects of Epistemological Beliefs; Stahl & Bromme, in press; appendix C3) had to be completed in reference to the domain of genetics and thus was used to measure domain-dependent epistemological beliefs. This instrument includes 24 pairs of connotative adjectives to measure the dimensions *texture* (structure and accuracy of knowledge; sample item: "structured – unstructured") and *variability* (stability and dynamics of knowledge; sample item: "dynamic – static"). All statements and adjectives in both instruments were rated on 7-point scales.

Because of the revisions of the WKI, an exploratory factor analysis was computed even with this small sample size. Although it was not possible to replicate the original factor

structure, a meaningful solution was obtained: The final solution explains 39 % variance and encompasses the factors *WKI-simplicity* (9 items, Cronbach's  $\alpha = .69$ ) and *WKI-certainty* (5 items, Cronbach's  $\alpha = .73$ ). The factor *WKI-simplicity* measures whether students assume that knowledge is an accumulation of facts versus a complex network of interrelated concepts (sample item: "Most words have one clear meaning."). The factor *WKI-certainty* refers to students' beliefs in absolute and exact versus tentative knowledge (sample item: "Today's facts may be tomorrow's fiction."). The final solution for the CAEB comprises the two original factors, *CAEB-texture* (11 items, Cronbach's  $\alpha = .81$ ) and *CAEB-variability* (9 items, Cronbach's  $\alpha = .80$ ) and explains 41 % of variance. Thus, all subsequent statistical analyses will be conducted with the epistemological beliefs scales *WKI-simplicity*, *WKI-certainty*, *CAEB-texture* and *CAEB-variability* as predictor variables or covariates.

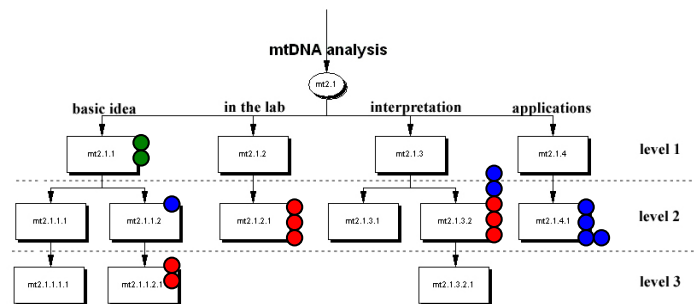


Figure 3.5-1: Visualization of the hypertext chapter about mtDNA analysis used in study III. The fourteen nodes are arranged on three hierarchical levels ("level 1 – 3") and belong to four topics ("basic idea", execution "in the lab", "interpretation", and practical "applications"). Links to "biological background" nodes are indicated by green circles, to "examples" by blue circles and to "problems" by red circles.

### *Hypertext on mtDNA Analysis*

The hypertext that was used in this study is a prototypical chapter of entire hypertext on "genetic fingerprinting" (chapter 3.2; appendix C5). More specifically, this hypertext version encompasses an eight-node introduction to genetic fingerprinting that provides general background knowledge necessary to understand the topic (e.g., structure of DNA). All students were instructed to read this introduction first. The main part of the hypertext (thirty-one nodes) explains the topic of mtDNA analysis (Figure 3.5-1). The hierarchical part of this hypertext chapter encompasses 14 nodes. The top two levels encompass five introductory nodes, one overall introductory node for mtDNA analysis and four introductory nodes for the sub-topics "basic idea" of mtDNA analysis, execution of mtDNA analysis "in the lab", "interpretation" of mtDNA analysis results and practical "applications" of mtDNA analysis. These nodes are simple and serve as introductions ("level 1"). The "level 2" contains six nodes of intermediate complexity that elaborate information on the four sub-topics. The "level 3" contains three complex nodes with very detailed information

on certain aspects of these sub-topics. Besides this hierarchical structure, the hypertext contains seventeen nodes in the appendices that are linked thematically with the main text. They belong to the categories of “biological background” (2 nodes, for example about the phenomenon of heteroplasmy), “examples” (7 nodes, for example special applications in forensic cases) and “problems” (8 nodes, for example problems with interpreting the results of an mtDNA analysis). Those nodes differ in length as well as complexity.

To conclude, the subsequent statistical analyses will include text complexity as defined by the three hierarchical hypertext levels as independent variable. Note that students could access all hypertext nodes in any order. In all analyses however the order will be depicted in ascending complexity to enhance comprehensibility.

#### *Comprehensibility Ratings*

Students had to judge the comprehensibility of each hypertext node they processed during hypertext learning on a 7-point scale ranging from 1 = very comprehensible to 7 = very incomprehensible (appendix C10).

#### *Knowledge Tests*

Eight multiple-choice questions were developed with the help of a domain expert to test students' prior domain knowledge about molecular biology (Cronbach's  $\alpha = .90$ , sample item: “What is the meaning of the abbreviation PCR?”; appendix C4).

To measure how much students learned about the specific topic of mtDNA analysis, 15 node-specific multiple-choice questions were developed. Each of these questions pertained to a main concept explained in one of 15 randomly selected hypertext nodes (mtDNA analysis knowledge test, Cronbach's  $\alpha = .60$ , sample item: “What constitutes heteroplasmy of the mtDNA?”; appendix C11). These questions covered the content of all parts of this hypertext chapter equally and were used as an indicator of learning outcome.

### **3.5.2 Results**

In this study  $p < .05$  was defined as significant and  $p < .10$  as marginally significant. Two humanities students were excluded from all further analyses because they did not comply with the instruction (they did not read the 8-node introduction first). Thus, the final sample for all subsequent analyses consists of 49 students.

For each hypertext part (the three hierarchical levels and the three appendices) three dependent variables were computed: (1) *Average Processing Duration per Node* (APDN = total time spent in a hypertext part divided by number of processed nodes in that part), (2) *Percentage of Processed Nodes* (PPN = number of processed nodes per part divided by number of existing nodes in that part) and (3) *Average Comprehensibility Rating* (ACR).

To give a more qualitative idea of students' hypertext navigation, the descriptive values for all hypertext parts including the appendices are displayed in Table 9. Students on average spent 38 minutes in the main hypertext ( $SD = 8:44$  minutes). During this learning phase students processed on average 16 nodes ( $SD = 5.78$ ). That equals 52 % of existing nodes (PPN:  $SD = 18.65$  %). On average each single node was processed for 2:29 minutes (APDN:  $SD = 0:32$ ). Not all students navigated to all parts of the hypertext. While all 49 students processed nodes on "level 1", 47 processed nodes on "level 2" and only 43 students navigated to "level 3". The appendices were processed less often: Only 14 students navigated to the "biological background", 22 students navigated to the "examples" and 17 students navigated to the "problem" nodes.

Table 9: Means (M) and standard deviations (SD) of dependent variables for all hypertext parts

M		SD		M		SD	
<b>Level 1 (n = 49)</b>				<b>Biological background (n = 14)</b>			
APDN	01:46	00:38		APDN	02:09	00:43	
PPN	95,10	13,86		PPN	78,57	25,68	
ACR	1,41	,62		ACR	1,75	,89	
<b>Level 2 (n = 47)</b>				<b>Examples (n = 22)</b>			
APDN	03:21	00:53		APDN	01:58	00:38	
PPN	87,23	20,33		PPN	48,70	30,75	
ACR	2,00	,76		ACR	1,96	1,17	
<b>Level 3 (n = 43)</b>				<b>Problems (n = 17)</b>			
APDN	03:35	01:27		APDN	01:50	00:45	
PPN	85,27	23,35		PPN	57,35	28,66	
ACR	2,59	1,17		ACR	2,03	,87	

APDN = *Average Processing Duration per Node*

PPN = *Percent of Processed Nodes*

ACR = *Average Comprehensibility Rating*

Generally, students navigated from introductory to more detailed hypertext nodes (from "level 1" through "level 2" to "level 3") and from left to right (Figure 3.5-2). Thus, a typical learning session started with exploring the "basic idea" behind mtDNA analysis including most of its subordinate nodes. Subsequently, a typical student navigated to the execution of mtDNA analysis "in the lab" with its subordinate node, followed by the nodes about the "interpretation" of results. Last, the practical "applications" were scrutinized. Furthermore, a typical student mostly utilized the hierarchical navigation tools ("parent", "child", and "siblings") and not the advanced tools to get to more distant hypertext nodes ("glossary",

“history”, “search” or “TOC” (table of content)). Additionally, a typical student rarely navigated to the “related information” nodes (“biological background”, “examples” or “problems”). However, a minority of students deviated from this pattern: For example, some students started with getting an overview and read the introductory nodes for all topics first. Subsequently, these students went back to the “basic idea” to explore the deeper level nodes for this topic. Other students differed in their use of navigational commands and extensively used non-hierarchical commands. Furthermore, some students also extensively used the related information nodes from the appendices “biological background”, “examples”, and “problems”.

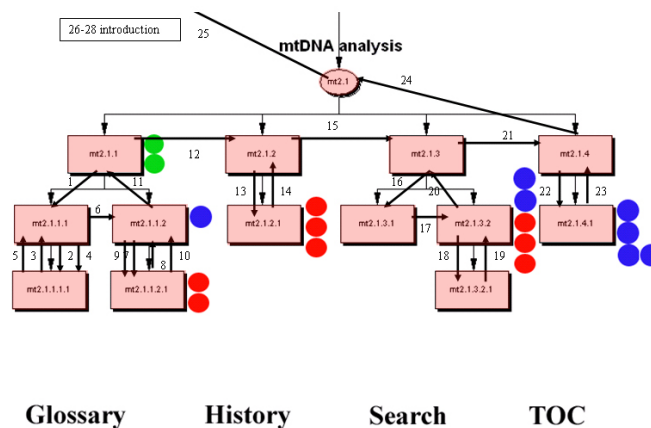


Figure 3.5-2: A typical navigation path through the hypertext chapter about mtDNA analysis (student AKMI29). Links to “biological background” nodes are indicated by green circles, “examples” by blue circles and “problems” by red circles. All visited hypertext nodes are colored in pale red (in this case all hierarchical nodes and none of the related nodes from the appendices).

Besides these variables capturing the learning process, the learning outcome was also considered as dependent variable. In this study the students had the free choice where to navigate in the hypertext to “learn as much as possible” about the topic mtDNA analysis. Nonetheless, all students had to answer the same multiple choice questions about mtDNA analysis. Therefore, not only the total test score of the knowledge test was computed, but also two sub-scores: the percentage of correctly answered questions pertaining to processed nodes ( $PP = \textit{Previously Processed}$ ) shows how much students recall the information they had in fact read. Furthermore, the percentage of correctly answered questions pertaining to non-processed nodes was calculated ( $PNP = \textit{Previously Not Processed}$ ). On average students answered 58 % of the fifteen questions about mtDNA analysis correctly ( $M = 8.76$ ;  $SD = 2.43$ ). When they processed the corresponding node previously (PP), they answered 86 % ( $SD = 15.09$  %) correctly. When they did not process the associated node previously (PNP) they answered 26 % ( $SD = 18.33$  %) correctly.



### 3.5.2.1 Do Students Discriminate Between Texts of Different Complexity?

To answer the first research question, a within-subject repeated-measure MANOVA was computed across the three hierarchical hypertext levels with all three dependent variables. Only students who navigated to all three levels could be included ( $n = 43$ , Table 9).

Results indicate a highly significant multivariate main effect for the repeated-measure factor hierarchical hypertext levels ( $F(6,38) = 48.91, p < .001$ ) which was replicated univariately on each dependent variable (APDN:  $F(2,86) = 45.71, p < .001$ ; PPN:  $F(2,86) = 5.37, p = .006$ ; ACR:  $F(2,86) = 45.59, p < .001$ ). The descriptive data presented in Table 9 indicates that for all dependent variables students continuously adapted to the complexity of the presented hypertext nodes. While nodes on higher hierarchical levels were processed shorter (APDN), the processing time continuously increased for deeper-level nodes. While almost all nodes on top-levels were accessed, the *percentage of processed nodes* (PPN) continuously decreased for deeper-level nodes. While students judged nodes on higher hierarchical levels quite comprehensible (ACR), the reported comprehensibility significantly decreased with increasing node complexity (on deeper hierarchical levels).

To summarize the results with regard to research question one: Students' hypertext navigation (APDN and PPN) as well as their comprehensibility ratings (ACR) demonstrate significant *discrimination* between hypertext levels of different complexity.

### 3.5.2.2 Do Students Calibrate to Text Complexity?

To answer the second research question measures of *relative calibration* were utilized: Within-subject Goodman-Kruskal Gamma correlations ( $G$ ) between the dependent variables (APDN, PPN, and ACR) and the hierarchical hypertext levels were computed ( $n = 3$ , for 3 hypertext levels) and subsequently Z-transformed into calibration indices. These calibration indices were tested against zero and scrutinized with regard to their effect sizes.

Calibration indices for all dependent variables significantly differ from zero and correspond to correlations of at least large effect size ( $G > .50$ ). More specifically, the mean calibration index for the *average processing duration per node* (APDN) is 1.46 ( $SD = 1.56$ ) which corresponds to a correlation of  $G = .90$  and significantly differs from zero ( $t(45) = 6.35, p < .001$ ). This positive association between the hierarchical hypertext levels and APDN indicates that students process hypertext nodes on deeper hierarchical levels much longer. For the dependent variable *percentage of processed nodes* (PPN), the picture is reversed (mean calibration index =  $-.73, SD = 1.68$ ; corresponds to  $G = -.63$ ; significantly differs from zero:  $t(48) = -3.02, p = .004$ ). This negative association between the hierarchical hypertext levels and PPN indicates that students process fewer nodes on deeper hierarchical levels. The mean calibration index for the dependent variable *average comprehensibility rating* (ACR) is 2.31 ( $SD = 1.20$ ) which corresponds to a correlation of  $G = .98$  and significantly differs

from zero ( $t(43) = 12.72, p < .001$ ). This positive association indicates that students give higher values in their comprehensibility judgments (indicating less comprehensibility) for nodes on deeper hierarchical levels.

To summarize the results with regard to research question two: Students' hypertext navigation (APDN and PPN) as well as their comprehensibility ratings (ACR) demonstrate significant *calibration* with regard to *hypertext complexity*.

### 3.5.2.3 The Impact of Prior Domain Knowledge and Epistemological Beliefs

Before reporting the relation between the *internal conditions* – in this case, students' prior domain knowledge (biology students vs. humanities students) and students' epistemological beliefs (indicated by their values on the scales *WKI-simplicity*, *WKI-certainty*, *CAEB-texture*, and *CAEB-variability*) – and students' *discrimination* and *calibration*, some descriptive values for the epistemological beliefs factors will be covered and all relations between these predictor variables will be detailed.

Students in this study tended to believe more in simple knowledge, thus displaying a “naïve” view on *WKI-simplicity* ( $M = 4.64, SD = .79$ ; on a 7-point scale from 1 = complex to 7 = simple). Furthermore, students believed strongly in uncertain knowledge, thus demonstrating a “sophisticated” view on *WKI-certainty* ( $M = 5.87, SD = .68$ ; on a 7-point scale from 1 = certain to 7 = uncertain). Additionally, students tended to believe more in structured knowledge in molecular genetics, thus displaying a “naïve” epistemological view on *CAEB-texture* ( $M = 2.94, SD = .71$ ; on a 7-point scale from 1 = structured to 7 = unstructured). And students tended to believe more in variable knowledge in molecular genetics, thus displaying a “sophisticated” epistemological view on *CAEB-variability* ( $M = 4.76, SD = .84$ ; on a 7-point scale from 1 = static to 7 = variable).

Correlational analysis revealed that *WKI-simplicity* was significantly related to *CAEB-texture* ( $r = -.29, p = .042$ ): Students who believed in simple knowledge in general also believed in structured knowledge in genetics. Furthermore, the correlation between the two domain-dependent factors of the CAEB, *CAEB-texture* and *CAEB-variability*, was significant ( $r = .43; p = .002$ ): Students who believed in unstructured knowledge in genetics also believed in relative knowledge in genetics. To also account for prior knowledge differences, epistemological beliefs on all scales were compared between biology and humanities students. Only one marginally significant difference was detected: Biology students tended to believe more in structured knowledge than humanities students (*CAEB-texture*:  $t(49) = 1.78, p = .081$ ; biology students:  $M = 2.77, SD = .68$ ; humanities students:  $M = 3.12, SD = .71$ ; all ratings on a 7-point scale from 1 = unstructured to 7 = structured).

*Are students' prior domain knowledge and their epistemological beliefs associated with their discrimination between texts of different complexity?*

To answer this research question, students' prior domain knowledge was included as a dichotomous factor (biology students vs. humanities students) and students' epistemological beliefs were included as covariates (*WKI-simplicity*, *WKI-certainty*, *CAEB-texture*, and *CAEB-variability*) in the within-subject repeated-measure MANOVA across the three hierarchical hypertext levels<sup>8</sup>. To visualize significant results, *calibration* graphs were created. For the epistemological beliefs scales, this visualization was achieved by median-split.

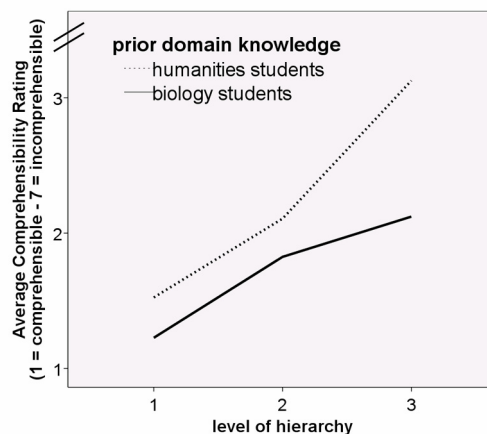


Figure 3.5-3: Calibration graph depicting students' *average comprehensibility ratings* (ACR) as a function of text complexity (hypertext levels) and prior domain knowledge (biology students vs. humanities students).

With regard to the effects of *prior domain knowledge*, the MANCOVA results revealed a significant multivariate main effect ( $F(3,35) = 3.47, p = .026$ ). This effect was corroborated univariately only on one dependent variable (ACR:  $F(1,37) = 7.51, p = .009$ , Figure 3.5-3): Biology students judged all nodes more comprehensible than humanities students. This effect was supported by correlational results: More prior knowledge (molecular knowledge test score) was associated with judging nodes more comprehensible on all three levels (Table 10). Furthermore, a significant univariate interaction between level of hierarchy and prior knowledge was detected for *average comprehensibility ratings* (ACR): biology students not only judged nodes on all levels more comprehensible (see above), but this effect became also more pronounced on deeper hierarchical levels ( $F(2,74) = 4.83, p = .011$ ; Figure 3.5-3). Additionally, results of the correlations demonstrated that more prior domain

<sup>8</sup> All assumptions of this analysis were tested. More specifically, the adequacy of the sample size for detecting medium effects was assured (Stevens, 1996) as well as an adequate number of covariates (Huitema, 1980). Furthermore, the normal distribution of covariates and dependent variables, the independence of covariates and treatment, the homogeneity of variance, and the homogeneity of hyperplanes was tested. Although small violations occurred, the results can still be considered valid (Grimm & Yarnold, 1995; Hair, Anderson, Tatham, & Black, 1998; Huitma, 1980; Tabachnick & Fidell, 2007).

knowledge was also associated with a lower *average processing duration of nodes* (APDN) on “level 1” and “level 2”, with higher *percentage of processed nodes* (PPN) on “level 2” and “level 3”, and with higher percentage of processed nodes in the “problem” nodes (Table 10).

With regard to the effects of *epistemological beliefs*, MANCOVA results revealed a significant multivariate main effect for the epistemological belief factor *WKI-simplicity* ( $F(3,35) = 4.43, p = .010$ ) and a marginally significant multivariate interaction between *WKI-simplicity* and level of hierarchy ( $F(6,32) = 2.38, p = .052$ ). The main effect was corroborated univariately on all three dependent variables (APDN:  $F(1,37) = 7.11, p = .011$ ; PPN:  $F(1,37) = 4.09, p = .051$ ; ACR:  $F(1,37) = 6.53, p = .015$ ). Students who believed in complex knowledge in general (“sophisticated” view on *WKI-simplicity*) processed single nodes significantly shorter (APDN), processed a marginally significant larger percentage of nodes (PPN) and judged nodes to be significantly less comprehensible (ACR) than their “naïve” counterparts who believed in simple knowledge. These results are supported by correlational results: The belief in complex knowledge (*WKI-simplicity*) was correlated with less comprehensible node evaluation on “level 2” and “level 3” (Table 10). The multivariate interaction effect of *WKI-simplicity* was univariately only replicated significantly on one dependent variable (APDN:  $F(2,74) = 6.94, p = .002$ ): Students who believed in complex knowledge (*WKI-simplicity*) not only processed nodes on all levels shorter (see above) but this difference became also more pronounced on deeper hierarchical levels (to visualize this effect, the epistemological belief factor *WKI-simplicity* was median-split, Figure 3.5-4, left). Consistently, correlational results also revealed a significant negative relation between the belief in complex knowledge (*WKI-simplicity*) and *average processing duration per node* (APDN) on “level 2” and “level 3” (Table 10).

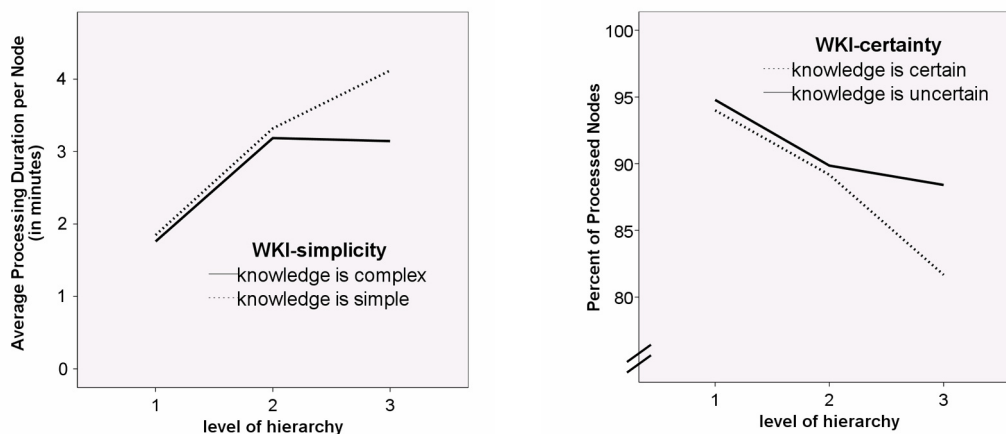


Figure 3.5-4: Calibration graphs depicting students *average processing duration per nodes* (APDN, left) and their *percentage of processed nodes* (PPN, right) as a function of text complexity (hypertext levels) and *WKI-simplicity* (median-split, left) or *WKI-certainty* (median-split, right).

MANCOVA results revealed no multivariate effect for the epistemological belief factor *WKI-certainty*. Still, a marginally significant interaction between *WKI-certainty* and level of hierarchy was detected univariately ( $F(2,74) = 2.77, p = .069$ ): Students who believed in uncertain knowledge in general (“sophisticated” view on *WKI-certainty*) tended to process more nodes (PPN) on deeper levels than their more “naïve” counterparts (to visualize this effect the epistemological belief factor *WKI-certainty* was median-split, Figure 3.5-4, right). This effect was corroborated by consistent correlational results: The belief in uncertain knowledge (*WKI-certainty*) was associated with a higher PPN on “level 3” (Table 10).

MANCOVA results revealed no significant effects for the CAEB. Nevertheless, the belief in unstructured knowledge in genetics (“sophisticated” view on *CAEB-texture*) was significantly correlated with judging “level 3” nodes less comprehensible (Table 10).

Table 10 also displays the correlations within the appendices: “Sophisticated” beliefs in unstructured knowledge (*CAEB-texture*) and complex knowledge (*WKI-simplicity*) were associated with judging “example” nodes less comprehensible. Furthermore, students who believed in unstructured knowledge (“sophisticated” view on *CAEB-texture*) processed fewer “problem” nodes (PPN) and students who believed in relative knowledge in genetics (“sophisticated” view on *CAEB-variability*) spent more time on “problem” nodes (APDN).

To summarize the results with regard to research question three and *discrimination: Prior domain knowledge* consistently impacts node evaluation (ACR), especially on deeper hierarchical hypertext levels: Biology students consider hypertext nodes more comprehensible than humanities students. *Epistemological beliefs* on the other hand mainly affect students’ hypertext navigation: “Sophisticated” epistemological beliefs are associated with processing single hypertext nodes shorter (APDN, *WKI-simplicity*) and with processing a higher percentage of hypertext nodes (PPN, *WKI-certainty*); both effects are especially pronounced for more complex texts on deeper hierarchical hypertext levels.

*Are students’ prior domain knowledge and their epistemological beliefs associated with their calibration to text complexity?*

To answer this research question, students’ epistemological beliefs on all scales (*WKI-simplicity*, *WKI-certainty*, *CAEB-texture*, and *CAEB-variability*) were correlated with their calibration indices for all dependent variables (APDN, PPN, and ACR). To also account for prior domain knowledge, the calibration indices of biology students and humanities students were compared via *t*-tests.

No impact of *prior domain knowledge* could be detected. Calibration indices of all dependent variables did not differ significantly between biology and humanities students. With regard to *epistemological beliefs*, only one significant correlation was found: “Naïve” beliefs in certain knowledge (*WKI-certainty*) were positively related to strong negative calibration ( $r = .34, p = .018$ ). This indicates that “naïve” students strongly differentiated between different hierarchical hypertext levels by processing a high percentage of nodes from top-levels and a

significantly lower percentage of nodes from deeper hierarchical levels. Students with “so-phisticated” beliefs on the other hand did not *calibrate* as much to hypertext levels.

Table 10: Correlations between covariates and dependent variables for all hypertext parts

	<i>WKI-simplicity</i>	<i>WKI-certainty</i>	<i>CAEB-texture</i>	<i>CAEB-variability</i>	Knowledge test score
<b>Level 1 (n = 49)</b>					
APDN	-.07	-.07	.03	-.19	-.30*
PPN	-.19	-.02	-.13	.09	.15
ACR	-.20	.03	.21	-.01	-.40**
<b>Level 2 (n = 47)</b>					
APDN	.09	-.08	.09	.06	-.32*
PPN	-.24	.19	-.18	.06	.33*
ACR	-.33*	-.14	.19	-.09	-.41**
<b>Level 3 (n = 43)</b>					
APDN	.42**	-.01	.06	.14	.00
PPN	-.13	.26 <sup>+</sup>	-.06	.14	.26 <sup>+</sup>
ACR	-.38*	-.01	.29 <sup>+</sup>	.09	-.51**
<b>Biological background (n = 14)</b>					
APDN	.19	-.14	-.05	-.15	-.12
PPN	.42	.28	.18	.28	.17
ACR	-.41	.08	.06	-.28	-.30
<b>Examples (n = 22)</b>					
APDN	.09	-.01	.25	.06	-.24
PPN	.00	-.12	-.31	-.32	.17
ACR	-.37 <sup>+</sup>	.05	.43*	.04	-.26
<b>Problems (n = 17)</b>					
APDN	.22	.13	.39	.56*	.09
PPN	.13	.37	-.57*	-.06	.49*
ACR	-.29	-.04	.17	-.03	-.15

APDN = *Average Processing Duration per Node*

PPN = *Percent of Processed Nodes*

ACR = *Average Comprehensibility Ratings*

\*\*p < .01; \* p < .05; + p < .10

To summarize the results with regard to research question three and *calibration*: *Prior domain knowledge* does not have an impact on calibration but “sophisticated” *epistemological beliefs* are negatively related to calibration.

### 3.5.2.4 Determinants of the Learning Outcome

The learning outcome can be defined by the number of correctly answered questions in the mtDNA knowledge test (total test score) or by two sub-scores: The number of correctly answered questions pertaining to previously processed nodes (PP = *Previously Processed*) or pertaining to previously non-processed nodes (PNP = *Previously Not Processed*). The impact of different quantitative variables (internal conditions, students’ calibration, and additional variables capturing their concurrent learning process) on these indicators of learning outcome will be scrutinized statistically. Additionally, a qualitative answer will be attempted by analyzing the learning process of the most and least successful students.

#### *Internal Conditions as Determinants of Learning Outcome*

An ANCOVA with prior domain knowledge (biology students vs. humanities students) as independent variable and the four epistemological beliefs scales (*WKI-simplicity*, *WKI-certainty*, *CAEB-texture*, and *CAEB-variability*) as covariates was computed for the *total test score*. Furthermore, a MANCOVA with the same variables was computed for the two sub-scores PP and PNP. The ANCOVA revealed a significant main effect of *prior domain knowledge* on the total test score ( $F(1,43) = 8.26, p = .006$ ). This effect was corroborated significantly within the MANCOVA multivariately ( $F(2,40) = 3.95, p = .027$ ) as well as univariately for PNP ( $F(1,41) = 6.69, p = .013$ ). The effect on PP was only marginally significant ( $F(1,41) = 3.15, p = .083$ ). In all cases, biology students scored higher on the mtDNA knowledge test. Significant correlations between the molecular biology test results also capturing students’ prior domain knowledge and the learning outcome scores support this pattern of effects (with total score:  $r = .56, p < .001$ ; with PP:  $r = .27, p = .065$ ; with PNP:  $r = .35, p = .018$ ). With regard to *epistemological beliefs*, the ANCOVA also revealed a marginally significant main effect of *WKI-simplicity* on the total test score ( $F(1,43) = 3.43, p = .071$ ): Students who believed in complex knowledge (“sophisticated” view) scored higher on the mtDNA knowledge test.

#### *Calibration Indices as Determinants of Learning Outcome*

Calibration indices of APDN, PPN, and ACR were correlated with the indices of learning outcome (total test score, PP, and PNP). The calibration indices were not significantly associated with any of the indices of learning outcome. For example, the correlations with

the *total test score* were nonsignificant (calibration of APDN:  $r = .17, p = .270$ ; calibration of PPN:  $r = .14, p = .348$ ; calibration of ACR:  $r = -.22, p = .16$ ).

#### *Further Learning Process Variables as Determinants of Learning Outcome*

Further variables capturing students' concurrent learning process (APDN, PPN, ACR; overall but also for different hypertext parts) were significantly correlated with the indices of learning outcome. As an example, the correlations with the *total test score* will be detailed: Across all hypertext parts, long processing of single nodes was significantly associated with fewer correctly answered questions (APDN:  $r = -.33, p = .023, n = 49$ ). This was especially true for hypertext nodes on "level 1" ( $r = -.37, p = .047, n = 49$ ), hypertext nodes on "level 2" ( $r = -.37, p = .010, n = 47$ ) and nodes within the "examples" ( $r = -.41, p = .061, n = 22$ ). Furthermore, across all hypertext parts, a higher *percentage of processed nodes* was significantly associated with more correctly answered questions (PPN:  $r = .59, p < .001, n = 49$ ). This was especially true for hypertext nodes on "level 1" ( $r = .29, p = .047, n = 49$ ), hypertext nodes on "level 2" ( $r = .44, p = .002, n = 47$ ) and nodes concerned with "problems" ( $r = .63, p = .007, n = 17$ ). Additionally, across all hypertext parts, judging nodes less comprehensible was significantly associated with answering fewer questions correctly (ACR:  $r = -.28, p = .049, n = 49$ ). This was especially true for hypertext nodes on "level 2" ( $r = -.41, p = .004, n = 47$ ) and those containing "examples" ( $r = -.52, p = .012, n = 22$ ).

#### *Qualitative Analysis of the Most and the Least Successful Student*

Success can be defined at least with regard to the total test score or with regard to PP (*Previously Processed*). The patterns for these two indicators of success partly differ, especially for the least successful students. Consider the example of student GHAW18. This student started out with extremely low prior knowledge (0 points on the molecular biology knowledge test). Thus, she was very slow in processing the hypertext and only managed to access 8 out of 32 nodes. Thus, her total score on the mtDNA knowledge test was quite low with only 4 points. But she managed to answer 60 % of the PP questions correctly. Thus, she was moderately successful on that score. For most students however, both indicators of success are fairly congruent. Two typical examples, one of a successful student and one of a less successful student will be described contrastively in further detail.

Student MJMM30 is a humanities student with moderate background knowledge in molecular biology (4 points in the prior knowledge test). With regard to the mtDNA knowledge test indicating learning outcome, MJMM30 only scored 3 points (38 % PP). This was the least successful students with regard to the total test score and with regard to PP.

As a more successful example, consider student IHMW06. She also is a humanities student with a good background knowledge in molecular biology (7 points in the prior knowledge test). With regard to the mtDNA knowledge test indicating learning outcome, IHMW06 scored 14 points (100 % PP). The only question she was not able to answer cor-



rectly referred to specific example that she had not read yet. This was the most successful student with regard to the total test score and one of 17 students with a perfect PP score.

What are the relevant similarities and differences between these two students (Why did they perform so differently)? In order to answer these questions, consider these students' navigation paths (Figure 3.5-5 and Figure 3.5-6). Both students used a similar number of navigational moves: MJMM30 made 62 moves and IHMW06 made 61 moves. Furthermore, both students frequently re-accessed previously processed hypertext nodes. Additionally, both students seem to follow the same typical navigation strategy: First these students explored the “basic idea” of mtDNA analysis, reading introductory nodes first, followed by more specific deeper-level nodes, then they accessed the nodes about the execution of mtDNA “in the lab”, moving on to the “interpretation” of the obtained results and the practical “applications”. Additionally, both students were well-calibrated with regard to their comprehensibility judgments: They consistently considered deeper level hypertext nodes more incomprehensible than top level nodes.

Students differed, however, on the following variables: First, the least successful student MJMM30 accessed 13 relevant hypertext nodes (plus the “history” window and one organizational node), while the most successful student IHMW06 accessed 26 relevant hypertext nodes (plus one organizational node). Note though, that a smaller number of accessed nodes did not automatically lead to a better recall performance: MJMM30 answered only 38 % of the PP questions correctly, while IHMW06 managed to answer 100 % of the PP questions correctly. Second with regard to the content, the most marked difference between these two students concerns the “related information” nodes. The most successful student IHMW06 accessed a high number of “related information” nodes (“biological background”, “examples”, and “problems”). These nodes were immediately accessed – and sometimes re-accessed – from their “base page” (the page they were linked to). Therefore, IHMW06 might have been able to successfully integrate the content of these “related information” nodes into her overall representation of the hypertext's content, thus deriving a more elaborate and complete situational model of the topic mtDNA analysis. The least successful student MJMM30 on the other hand did not access any such node. Therefore, her internal model of mtDNA analysis necessarily remained incomplete and fragmentary. Third, IHMW06 seemed to have navigated more systematically and parsimonious than MJMM30. More specifically, IHMW06 accessed almost all necessary nodes and few unnecessary nodes. Most of her re-accesses of nodes seem to be systematic, probably to compare content, thus even these re-accesses might not be considered unnecessary moves. MJMM30 on the other hand accessed only half as many necessary nodes and made many unnecessary – and sometimes seemingly erratic – navigational moves by re-accessing the same nodes time and again. It is unlikely that she systematically re-read the content that many times (e.g., the introductory node mt.2.1 was accessed more than 10 times but only encompasses a few sentences). Fourth, the most successful student IHMW06 was well-

calibrated with regard to APDN (i.e., processed deeper level nodes longer than top level nodes) while the least successful student MJMM30 was negatively calibrated with regard to this variable (i.e., processed top level nodes longer than deeper level nodes). Therefore, the most successful student seemed to develop an elaborate representation of deep level nodes' content while the least successful student seemed to develop only a surface representation of these nodes. Even though only these two examples were contrasted, these differences are typical of less successful and successful students. From further examples it can be deducted, however, that more successful student might also read the hypertext selectively and might skip whole chapters or sub-chapters if they possessed sufficient prior domain knowledge. Furthermore, some of these successful students also successfully employed advanced navigational commands such as the "TOC" (table of content) or the "history".

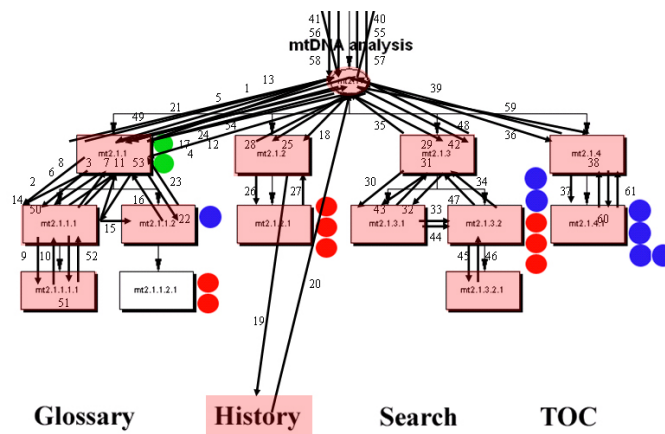


Figure 3.5-5: Navigation path of student MJMM30. All visited hypertext nodes are colored in pale red (in this case almost all hierarchical nodes and none of the related nodes from the appendices).

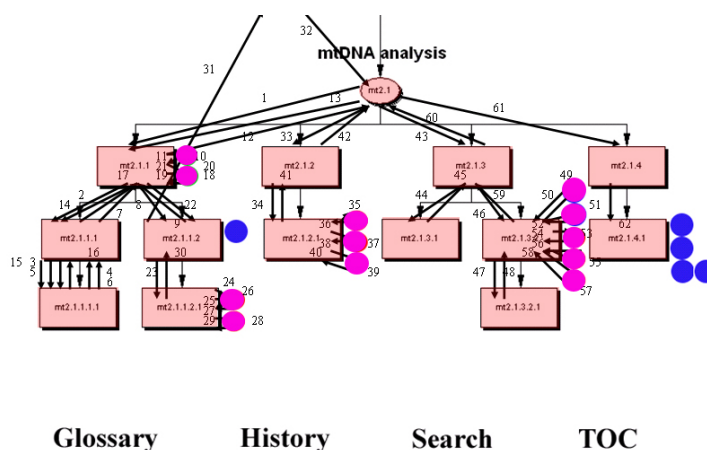


Figure 3.5-6: Navigation path of student IHMW06. All visited hypertext nodes are colored in pale red (the hierarchical nodes) or pink (the related information nodes).

### *Short Summary of Results*

The results of the quantitative analyses indicate that high prior domain knowledge and more “sophisticated” epistemological beliefs are positively related to learning outcome. Additionally, processing single nodes on top levels shorter (APDN), processing more nodes (PPN), and judging texts more comprehensible (ACR) are also positively related to learning outcome. However, calibration indices are no significant predictors. Additionally, the qualitative results comparing the most and least successful students indicate that successful students access a higher number of relevant nodes, especially a higher number of “related information” nodes, only demonstrate a limited number of re-accesses (small number of irrelevant nodes or moves), and thus overall demonstrate a more systematic and parsimonious navigation strategy.

### **3.5.3 Local Discussion**

The main goal of this study was to explore students’ *discrimination* between and *calibration* to *text complexity* (different hypertext levels). Students navigated through a small hypertext (31 nodes) with the rather open goal to “learn as much as possible” about mtDNA analysis. Keep in mind that all presented results should be interpreted with due caution: This study was designed as a first exploration of these research questions with a small sample size and furthermore effects were interpreted starting at  $\alpha < .10$ .

#### **3.5.3.1 Discrimination and Calibration**

Students consistently and significantly demonstrated their ability to adapt their learning process to the complexity of the hierarchical hypertext levels: The analyses concerning students’ *discrimination* and *calibration* consistently indicate positive results.

More specifically, analyses with regard to students’ *discrimination* indicate significant differences in students’ hypertext navigation (APDN and PPN) and in their comprehensibility ratings (ACR) between different hypertext levels. With regard to students’ *calibration* to text complexity (hierarchical hypertext levels), students demonstrate good *relative calibration*: All calibration indices (APDN, PPN, and ACR) differ significantly from zero and correspond to correlations of at least large effect size. All analyses consistently indicate that students processed nodes on deeper hierarchical levels longer (APDN), processed a smaller percentage of deeper-level nodes (PPN), and considered nodes on deeper hierarchical levels less comprehensible (ACR). Thus, students metacognitively monitored text complexity and metacognitively regulated their hypertext navigation accordingly. The success of this adap-

tive behavior is underlined by students' ability to answer the majority of knowledge questions about mtDNA analysis correctly.

These results are consistent with the COPES-model (Winne & Hadwin, 1998; chapter 2.1) that assumes that students systematically adapt their learning process to external conditions. Additionally, these results are also consistent with those of previous empirical studies about text complexity (chapter 2.4) indicating that students process differently complex texts differently (Maki, Foley, Kajer, Thompson, & Willert, 1990; McNamara, Kintsch, Songer, & Kintsch, 1996; Schommer & Surber, 1986; Weaver & Bryant, 1995). However, none of these empirical studies investigated students' concurrent text processing directly; instead they were inferred from students' learning outcome and calibration of comprehension monitoring. Furthermore, a multitude of different operationalisations of text complexity was employed: for example, intact text vs. a text with deleted letters (Maki, Foley, Kajer, Thompson, & Willert, 1990), texts varying in their Flesch scores (Weaver & Bryant, 1995) or texts varying in coherence (McNamara, Kintsch, Songer, & Kintsch, 1996).

These examples highlight the additional benefit of this study: First, text complexity was tested empirically a priori by subjective student judgments, and second, not only learning outcome measures were considered, but the primary focus was the concurrent process of text processing.

### ***3.5.3.2 The Impact of Prior Domain Knowledge and Epistemological Beliefs***

With regard to the impact of *internal conditions* on students' *discrimination*, students' prior domain knowledge (biology students vs. humanities students) as well as their epistemological beliefs (captured by the scales *WKI-simplicity*, *WKI-certainty*, *CAEB-texture*, and *CAEB-variability*) demonstrated significant impact. With regard to the impact of internal conditions on students' *calibration* only one significant effect for *WKI-certainty* could be detected. Additionally, both internal conditions impacted the *learning outcome*.

#### *Prior Domain Knowledge*

Prior domain knowledge was primarily related to students' comprehensibility ratings and their learning outcome but had almost no impact on students' navigation: The discrimination analysis revealed that biology students judged nodes across all hypertext levels more comprehensible than humanities students and this effect became more pronounced on deeper hierarchical levels (ACR). Most likely, biology students were familiar with some of the facts explained in the hypertext. Thus, the nodes were subjectively simpler to comprehend which influenced their text evaluations. An additional correlational analysis revealed that more prior domain knowledge was also associated with shorter processing of nodes on "level 1" and "level 2" (APDN), and processing of more nodes on "level 3" (PPN). Based on their prior domain knowledge students might have been faster to comprehend the

nodes' content (at least on top levels, APDN) and thus were able to access more nodes (at least on deeper levels, PPN). Finally, the analysis regarding the determinants of the learning outcome revealed that biology students were better able to answer all questions about mtDNA analysis. Nonetheless, these effects might be more selective: For questions pertaining to processed nodes the difference between biology and humanities students was only marginally significant (PP:  $p = .083$ ) whereas the difference was highly significant if questions pertaining to non-processed nodes were considered (PNP:  $p = .013$ ). Thus, prior domain knowledge seems to be especially helpful in compensating for non-processed material. Biology students were probably able to infer the right answers from their prior knowledge and already processed nodes.

These effects are consistent with the predictions of the COPEs-model (Winne & Hadwin, 1998; chapter 2.1): Prior domain knowledge impacted students' standards (indirectly captured by ACR), their operations (directly captured by APDN and PPN), and their learning outcome. Additionally, these results are also consistent with those of previous empirical studies about prior domain knowledge (chapter 2.5). For example, the fact that prior domain knowledge was not equally important for texts of different complexity was also demonstrated by Salmerón, Kintsch, and Canas (2006) who manipulated text coherence which can be considered an indicator of complexity and by McDonald and Stevenson (1998) who manipulated the complexity of a hypertext by linking it linearly, hierarchical, and in a mixed way. In both examples, prior domain knowledge elicited more effects with more complex (hyper)texts (also Calsir & Gurel; 2003; Möller & Müller-Kalthoff, 2000; McNamara, Kintsch, Songer, & Kintsch, 1996; Potelle & Rouet, 2003). Furthermore, Rouet, Favart, Britt, and Perfetti (1997) also report consistent results with regard to the differential effects of prior domain knowledge: In their study, it did not impact the employed learning strategies (comparable to the small impact of prior domain knowledge on operations in this study; APDN and PPN) but significantly impacted students' document evaluation (comparable to the significant impact of prior domain knowledge on students' comprehensibility ratings; ACR). Additionally, most empirical studies consistently indicate the benefit of expertise for the learning outcome (Ford & Chen, 2000; Lawless, Brown, Mills, & Mayall, 2003; Lind & Sandmann, 2003).

### *Epistemological Beliefs*

Epistemological beliefs had a significant impact on students' comprehensibility ratings, their navigation, and their learning outcome.

With regard to their comprehensibility ratings, more "sophisticated" beliefs (*WKI-simplicity* and *CAEB-texture*) were consistently positively related to considering hypertext nodes less comprehensible (ACR). Most likely, students with "sophisticated" epistemological beliefs judged comprehensibility in terms of an elaborate critical internal standard (deep

understanding) whereas more “naïve” students might have been satisfied with understanding single facts without considering their interrelations (superficial standard).

Within the main *hierarchical part of the hypertext* “sophisticated” beliefs were consistently associated with a selective reading strategy (Reynolds, 1992; Hidi, 1995): Students with more “sophisticated” epistemological beliefs read each single node selectively, probably concentrating on the most relevant content. For example, within the discrimination analyses “sophisticated” beliefs with regard to *WKI-simplicity* were associated with processing nodes shorter (APDN). Therefore, more “sophisticated” students were able to get a more comprehensive overview of the whole hypertext by processing more nodes. For example, within the discrimination analyses “sophisticated” beliefs with regard to *WKI-simplicity* and *WKI-certainty* were associated with a high number of processed nodes (PPN), for the latter epistemological beliefs scale especially on deeper hierarchical levels. Students with more “naïve” epistemological beliefs on the other hand might have been seduced by extraneous irrelevant details thus spending more time on each node (APDN) and therefore they might have been unable to access as many nodes (PPN). Two conclusions can be drawn: First, “sophisticated” epistemological beliefs were consistently associated with the most beneficial learning strategy. This is explicitly corroborated by the results regarding the determinants of the learning outcome that show that students believing that knowledge is complex (“sophisticated” view on *WKI-simplicity*) demonstrated higher total *learning outcomes* than students who believed in simple knowledge. And second, these arguments might also explain the seemingly counterintuitive calibration effect that more “sophisticated” beliefs with regard to *WKI-certainty* were associated with less calibration with regard to the percentage of processed nodes (PPN): Strong calibration necessarily demands that fewer hypertext nodes on deeper hierarchical levels are processed. More “sophisticated” students on the other hand processed almost all nodes – even on deeper hierarchical levels. Consequently, the counterintuitive calibration effect might be due to a ceiling effect: “Sophisticated” students’ content coverage (PPN) was too high for all hypertext levels.

Within the *appendices* of the hypertext on the other hand “sophisticated” beliefs were consistently associated with a deep elaboration strategy: Students with more “sophisticated” epistemological beliefs read each node longer (correlation between *CAEB-variability* and APDN), probably to scrutinize information about potential “problems” that may assist the critical evaluation of the factual information presented in the main part of the hypertext, and thus were unable to access a high number of these nodes (correlation between *CAEB-texture* and PPN). However, this strategy was not the best to score high on the knowledge test (see determinants of the learning outcome): With regard to the appendices, “sophisticated” epistemological beliefs were associated with a detrimental learning strategy.

Considering the whole pattern of effects they are mostly consistent with the COPES-model (Winne & Hadwin, 1998; chapter 2.1) because they indicate that epistemological beliefs have in fact an impact on students’ standards (indirectly captured by ACR), their

operations (directly captured by APDN and PPN), and their learning outcome. Additionally, these results are also mostly consistent with those of previous empirical studies about epistemological beliefs (chapter 2.6). For example, considering hypertexts less comprehensible is consistent with setting deeper internal standards which is also indicated by other empirical studies: “Sophisticated” students apply deeper criteria for comprehension (Ryan, 1984), are more likely to set mastery goals (Bråten and Strømsø, 2004), and apply more strategies to develop awareness (Wood & Kardash, 2002). Furthermore, epistemological beliefs are often associated with beneficial learning strategies (Bendixen & Hartley, 2003; Jacobson & Spiro, 1995). For example, Bartholomé, Stahl, Pieschl, and Bromme (2006) demonstrated that students with more “sophisticated” beliefs accessed more context-sensitive help nodes during plant identification. This access of more information could be interpreted as indicating the same underlying strategy of getting a thorough overview employed by the “sophisticated” students in this study. Additionally, most empirical studies indicate the benefit of “sophisticated” beliefs for the learning outcome that was also detected here (Schommer, 1990; Schommer, Crouse, & Rhodes, 1992). However, also note some discrepancies with theoretical assumptions: The COPES-model would have assumed a beneficial effect of “sophisticated” epistemological beliefs for all components of the learning process while in this study “sophisticated” beliefs were also associated with applying a detrimental strategy in the appendices and with less calibration. Note that other empirical studies also found such counterintuitive effects of epistemological beliefs: For example, “sophisticated” beliefs were associated with reporting less deep processing strategies (Cano, 2005), with reporting more surface strategies (Dahl, Bals, & Turi, 2005), and with inaccurate text processing (Kardash & Howell, 2000).

### ***3.5.3.3 Determinants of the Learning Outcome***

Students’ prior domain knowledge, their epistemological beliefs, their hypertext navigation strategies and their comprehensibility judgments significantly predicted the learning outcome. As elaborated in the previous section, more prior domain knowledge as well as more “sophisticated” epistemological beliefs were positively related to the learning outcome.

Furthermore, considering hypertext nodes more comprehensible (ACR) was also positively related to learning outcome. Additionally, a *selective reading strategy* was beneficial: Spending much time on single nodes (APDN) was negatively related to learning outcome, especially for nodes from “level 1”, “level 2” and from the “examples”. This makes sense considering that top-level nodes contain more introductory material that is elaborated on a more detailed level on deep-level nodes and considering that the content of the “examples” just applies the content of the main hypertext to specific cases. Thus, processing these nodes very deeply might not advance students’ understanding of the subject matter. On the other hand processing a high percentage of nodes (PPN) was positively related to learning

outcome, especially for nodes from “level 1”, “level 2”, and from the “problems”. This signifies that reading top-level nodes at least superficially might be necessary for understanding the hierarchically subordinate nodes and that reading of the “problem” nodes significantly contributes to the overall understanding of the subject matter.

However, no relation between students’ *calibration* indices for the dependent variables (ACR, APDN and PPN) and their learning outcome could be detected even though the COPES-model (Winne & Hadwin, 1998; chapter 2.1) predicted that pronounced calibration should result in better learning outcome. It is unclear if this lack of effects might be due to the time constraints or the special demands of the knowledge test (see below).

#### **3.5.3.4 Open Issues and Implications**

This study was successful at demonstrating that students in fact do *discriminate* between texts of different complexity and *calibrate to text complexity* and that these processes are impacted by students’ *prior domain knowledge* and *their epistemological beliefs* in the *enactment* stages of learning (previous sections). However, some limitations have to be considered.

##### *Open Issues and Limitations*

First of all, some issues with regard to *hypertext navigation* need to be considered: Students could navigate through the hypertext in a self-determined way. Thus, students could consciously decide to skip some hypertext nodes and concentrate on others instead. This selection process is a defining characteristic of learning with complex information systems. Thus, it is very interesting to notice what kind of information is attended and what kind of information is ignored. Nonetheless, it reduces the sample size for analyses, especially with regard to the appendices that were rarely accessed in this study. Additionally, all students had the same amount of time for learning and therefore the percentage of processed nodes (PPN) and the average processing duration per node (APDN) were interdependent. Students could either concentrate on getting an overview by processing as many nodes as possible (high PPN) or concentrate on paying adequate attention to details by spending long time on each node (high APDN). Furthermore, so far only single unit measures describing the online hypertext navigation were included as dependent variables (see discussion of study II). These variables do not capture more detailed hypermedia navigation strategies. However, as there are only few studies investigating, for example, the impact of epistemological belief on *online* learning processes this can be considered an adequate first strategy.

Second, the validity of students’ *comprehensibility ratings* (ACR) could be questioned: It could be argued that students do not really rate the comprehensibility of a text but instead use the position of a specific hypertext node within the hierarchical structure of the hypertext as cue for their ratings. However, this argument can be refuted by having a more detailed look at the data: For almost all node combination counterexamples can be found



where the same student judged a node on a higher level less comprehensible than a node on a deeper level. Thus, students do not seem to have used this cue – at least not consistently and systematically. An additional issue with regard to comprehensibility ratings concerns their relation to knowledge: As indicated by the consistent associations between prior domain knowledge and comprehensibility ratings in this study (see below), it is unlikely that students' can "objectively" judge the text's comprehensibility. Rather text comprehension is always an interaction between objective text properties and subjective learner characteristics (chapter 2.4 on text complexity).

A third issue pertains to *prior domain knowledge*: This learner characteristic had surprising little impact on students' hypertext navigation (APDN and PPN) in comparison with its consistent and significant impact on students' comprehensibility ratings (ACR) and their learning outcomes. A potential explanation concerns the domain-specificity versus domain-generalness of expertise: Hypertext navigation might be more dependent on domain-general approaches to learning (e.g., students' metacognitive knowledge and strategies for adequately exploring new domains) than on prior domain knowledge for a specific topic thus explaining the lack of effects.

A related fourth open issue pertains to the question to which degree *epistemological beliefs* are domain related or independent from the domain they refer to. There is growing evidence that learners have general epistemological beliefs, as well as domain related beliefs (Buehl, Alexander & Murphy, 2002). But up to now it is unclear how such different levels might interact with each other. Due to this discussion one domain-general instrument (WKI) and one domain-dependent instrument (CAEB) were included in this study. Surprisingly, the students' more general beliefs, especially concerning the dimension of *WKI-simplicity*, had the strongest impact on the discrimination and calibration processes within the hierarchical hypertext and on the learning outcome. Within the appendices the domain-related epistemological belief factors captured by the CAEB had the strongest impact whereas the domain general factors of the WKI almost had no impact. Therefore, it could be assumed that the domain-specific epistemological beliefs only become relevant for processes such as critical evaluation (based on the "problem" nodes) or further elaboration (based on "examples"). For more general navigational behavior – that might be less dependent on a specific topic but more on how to handle the structure of the hypertext – the domain general beliefs appear to be more relevant.

A fifth issue concerns the *knowledge test about mtDNA analysis* administered as measure of learning outcome. In this study a selective reading strategy was most beneficial for good scores on this learning outcome measure. However, this might only be the case for such a multiple-choice knowledge test focusing on recognition of main concepts. In other contexts it might have been more beneficial to employ other strategies. For example, if a learning outcome test required deep conceptual understanding it might have been better to employ deep elaboration strategies. Thus, it would be interesting to explore this issue and in-

investigate students' strategy use under ideal conditions (e.g., when they aim at deep understanding, are intrinsically motivated, and have as much time as they want). It seems safe to assume that such ideal conditions would elicit different strategies. This is supported by data indicating that students in this study did not reach a "ceiling" yet: Only few students processed the detailed content given in the deep-level nodes ("level 3") with due attention and only few students accessed further "examples" or "problems" to better comprehend and evaluate the given content. Note however, that this does not imply that *calibration* to text complexity had to be detrimental per se for the task to "learn as much as possible" and the corresponding knowledge test: In fact, a fair number of students with very good learning outcome are well-calibrated, especially with regard to their time spent on single nodes (APDN): They significantly processed deep-level nodes longer than top-level nodes. Thus, it would have been possible theoretically to find a positive correlation between calibration and learning outcome – which was not the case in this study.

### *Implications*

Because of the above mentioned open issues only tentative implications can be drawn. First of all, the results show that students do metacognitively calibrate their learning process to text complexity (hierarchical hypertext level). Furthermore, prior domain knowledge and epistemological beliefs most strongly impact the processing of complex, deep-level nodes or "problem" nodes. These differential effects of the internal conditions imply that the use of complex information sources or information sources that allow for critical evaluation requires special attention in educational settings. One way to deal with these differences in processing is to provide adequate scaffolding, for example by eliciting an adequate understanding of the task or by stimulating adequate goals (deep elaboration).

Consider, for example, that the task to "learn as much as possible" could be interpreted in different ways. One perspective would be to try to get an overview of the whole topic by browsing as many nodes as possible. Students with such a perspective would regulate their learning process with the goal of getting as complete and deep an overview as possible in the short time given by the experimenters ("hard" constraint). Another interpretation would be to memorize each detail presented. This interpretation would neglect to control for limited time and just consider the targeted level of understanding (i.e., memorization of details). In this study, employing a selective reading strategy in the main hierarchical part of the hypertext consistent with the first interpretation proved more beneficial for the learning outcome than a deep elaboration strategy consistent with the second interpretation. Other interpretations of this task are feasible.

To conclude, scaffolds for task interpretation might – at least for students with more "naïve" epistemological beliefs or low prior domain knowledge – ameliorate students' spontaneous processing of such complex or critical information sources.

## 4 General Discussion

“... I employed a strategy learned in Geography: you just pick some position that you like – or the same in German with interpretations – and you devalue the other information and collect your own information which are congruent with your hypothesis.”

(student MHJH20, study II)

The theoretical contributions of this thesis will be elaborated by relating the empirical results of this thesis systematically to the COPEs-model (Winne & Hadwin, 1998) that was used as a general framework of experimental design. First, the consistency of these findings with the core assumptions of the COPEs-model will be reviewed (chapter 4.1). Second, as one of the most important research questions students' adaptation to external conditions will be discussed in detail (*processes of calibration*). All discrimination and calibration results will be examined, especially the question if the transfer of methodology from the traditional calibration paradigm was successful (chapter 4.2). Additionally, the differential impact of task complexity and text complexity will be illuminated (chapter 4.3). Third, as another important research question the impact of students' prior domain knowledge and epistemological beliefs on these processes of adaptation will be scrutinized in detail (*conditions and processes of calibration*; chapter 4.4), especially the question if these conditions impact all stages of self-regulated learning alike (chapter 4.5). Fourth, the impact of these conditions and processes of adaptation on the learning outcome will be examined (“*To calibrate or not to calibrate?*”; chapter 4.6). Considering that the COPEs-model itself is not well-suited for the deduction of specific hypothesis because it is underspecified with regard to multiple issues (chapter 2.1), an adapted version will be proposed as a result of this review that is tailored to the specific empirical results from this series of studies (chapter 4.7). Note that these theoretical implications can not only be used to specify the COPEs-model but also contribute to the whole self-regulated learning research tradition.

The practical contributions of this thesis will be shortly discussed subsequently, especially the benefit of the epistemological sensitization implemented to elicit more “sophisticated” epistemological beliefs and deeper learning processes (chapter 4.8).

Last but not least potential future research will be highlighted; especially studies that help answer the remaining open questions of this thesis and studies that would help to generalize the results to other settings (chapter 4.9).

#### 4.1 Are the Empirical Results Consistent with the COPES-Model?

Because the COPES-model (Winne & Hadwin, 1998) is underspecified with regard to some issues it is hard to refute any part of it. Only if one of the following empirical results was found the validity of the COPES-model could be seriously doubted (P. Winne, personal communication, 13<sup>th</sup> September, 2006): (1) Students do not *metacognitively* regulate their behavior at all. (2) Students do not possess *internal standards* to compare ongoing processes with. (3) No *preparatory* stage of self-regulated learning could be detected. (4) Students do not attend to *external conditions* at all. (5) Students *internal conditions* have no impact at all.

All empirical results within this thesis are consistent with these postulations: (1) Students concurrent thoughts and their retrospective reports of their learning strategies consistently indicate that they *metacognitively* regulate their self-regulated learning process (e.g., study II). (2) The existence of *internal standards* is supported, for example by students' ability to report their internal standards (study I and study II). (3) The existence of a *preparatory* stage of learning is supported, for example, by students' concurrent thoughts, their retrospective self-reports about their learning processes (study II), and by their ability to give meaningful judgments without actually solving tasks (study I). (4) All empirical studies of this thesis demonstrate – as a central result – that students adapt their whole self-regulated learning process meaningfully to the *external conditions* task complexity and text complexity (study I – III). (5) Additionally, all empirical studies demonstrate – as a central result – that these adaptation processes are significantly impacted by the *internal conditions* prior domain knowledge and epistemological beliefs (study I – III).

#### 4.2 Do Learners Adapt their Learning to External Conditions?

Because the COPES-model (Winne & Hadwin, 1998; chapter 2.1) offered no detailed predictions with regard to this question, this general framework was supplemented with a more detailed conceptualization, an adequate terminology and a research methodology transferred from the *traditional calibration paradigm* (chapter 2.2). With this extended construct definition of “calibration” it was possible to empirically investigate students' processes of adaptation to external conditions in detail.

In the traditional calibration paradigm, calibration is diagnosed by comparing students' *metacognitive judgments* (e.g., regarding confidence) with *their performance* on a criterion test. Therefore, it is assumed to capture students' metacognitive monitoring. Based on a critical review of this paradigm two extensions of the construct “calibration” were proposed. (1) An extension of the criterion was suggested: Calibration should not only refer to a comparison between students' metacognitive judgments and their own performance (internal criterion), but also to a comparison between students metacognitive judgments and *external*

*criteria*. (2) An extension of metacognitive judgments was suggested: Calibration should not only refer to a comparison between students' metacognitive judgments and external criteria, but also to a comparison between *students' enacted metacognitive control strategies* and external criteria. With these extensions potential problems associated with the traditional calibration paradigm were alleviated and the fit between students' whole self-regulated learning process and external criteria could be investigated in detail with this methodology. Therefore, the central research question of this thesis could be re-conceptualized as a question of calibration: How exactly do students calibrate their metacognitive judgments and their enacted metacognitive control strategies to external criteria?

The traditional calibration paradigm diagnoses calibration with a specific methodology: *Discrimination* determines if students' metacognitive judgments differ significantly between situations. To determine the degree of correlation between metacognitive judgments and students' performance measures of *relative calibration* are used. To determine the absolute accuracy of students' judgments measures of *absolute calibration* are used. These relationships between students' metacognitive judgments and their performance can be visualized by *calibration graphs*. To apply this methodology to the central research question of this thesis, the relevant external criteria needed to be systematically manipulated. Based on a conceptual analysis of hypertext learning two external criteria (external conditions according to the COPES-model) were considered most relevant for this thesis: *task complexity* and *text complexity*. Task complexity was operationalised by the hierarchical categories of Bloom's revised Taxonomy (Anderson et al., 2001; chapter 2.3) and text complexity was operationalised by differently complex hierarchical hypertext levels (chapters 2.4 and 3.2).

#### *Study I: Calibration to Task Complexity in the Preparatory Stage*

Students were presented with six tasks of different complexity (external condition: *remember, understand, apply, analyze, evaluate, and create*) and had to give (metacognitive) judgments about important *conditions, operations, standards, and evaluations* for each task. This experimental design is similar to the one employed in studies within the traditional calibration paradigm because it also focuses on students' ability to monitor and give judgments accordingly. However, it differs significantly from the traditional calibration paradigm because students' judgments are not compared to their own performance (internal criterion) but with task complexity defined by Bloom's revised Taxonomy (external criterion). Results are consistently positive. They indicate highly significant *discrimination* for all dependent variables: Students' judgments differed significantly between the differently complex tasks. Results also indicate significant *relative calibration* for all dependent variables: Calibration indices differed significantly from zero and correspond mostly to correlations of large effect size:  $G = .47$  to  $G = .85$ . Additionally, *absolute calibration* could be determined for one item in this study: Students were able to classify about 50 % of the tasks in the correct categories of Bloom's revised Taxonomy, thus demonstrating good absolute accuracy.

*Study II: Calibration to Task Complexity in the Enactment Stage*

Students had to actively solve tasks of different complexity (external condition: *remember, remember, evaluate, understand, and remember*) with a hypertext on “genetic fingerprinting”. Simultaneously, their concurrent thoughts and their concurrent actions were captured. This experimental design constitutes an even further transfer of the methodology from the traditional calibration paradigm. Instead of capturing students’ metacognitive processes by metacognitive judgments, students’ whole self-regulated learning process was captured to also determine their metacognitive control strategies (concurrent thoughts and actions). Additionally, these processes are not compared with students own performance (internal criterion) but with task complexity defined by Bloom’s revised Taxonomy (external criterion). Results are consistently positive. They indicate highly significant *discrimination* for all dependent variables: Students’ concurrent thoughts and actions differed significantly between the differently complex tasks. Results also indicate significant *relative calibration* for all dependent variables: Calibration indices differed significantly from zero and correspond to correlations of large effect size:  $G = .88$  to  $G = .97$ .

*Study III: Calibration to Text Complexity in the Enactment Stage*

Students had to “learn as much as possible” about mtDNA analysis with a hypertext chapter about that topic that encompassed three hierarchical levels (external condition text complexity: “level 1”, “level 2”, and “level 3”). Simultaneously, their concurrent actions and their judgments regarding text comprehensibility were captured. This experimental design also constitutes a far transfer of the methodology of the traditional calibration paradigm. Instead of capturing students’ metacognitive processes only by metacognitive judgments, students’ whole self-regulated learning process was captured to also determine their metacognitive control strategies (concurrent actions). Additionally, these processes are not compared with students own performance (internal criterion) but with text complexity defined by the hierarchical hypertext level (external criterion). Results are consistently positive. They indicate highly significant *discrimination* for all dependent variables: Students’ concurrent actions and their comprehensibility judgments differed significantly between the hierarchical hypertext levels. Results also indicate significant *relative calibration* for all dependent variables: Calibration indices differed significantly from zero and correspond to correlations of large effect size:  $G = -.63$  to  $G = .98$ .

*Issues and Limitations*

First, one limitation of this transfer of methodology should be considered: In most studies of the traditional calibration paradigm it is possible to determine measures of *relative calibration* as well as measures of *absolute calibration* which is advocated (Schraw, 1995). However, this is not possible for the extended construct “calibration” as investigated in this series of

empirical studies. Only relative calibration can be diagnosed because no comprehensive prescriptive model of ideal self-regulated learning for different external condition exists. The empirical data allow for no conclusions about the absolute fit between learners' self-regulation and external conditions. Note though that even determining relative calibration is a valuable addition to research on self-regulated learning as such a systematic investigation of the *relationship* between students' self-regulated learning and potentially relevant external criteria was never attempted before with this methodology.

The second issue concerns *task complexity*: The detected results are consistent with other empirical studies indicating that learners process differently complex tasks differently (Klayman, 1985; Rouet, 2003; Rouet, Vidal-Abarca, Erboul, & Millogo, 2001; Veenman & Elshout, 1999; Winne & Jamieson-Noel, 2003). Scrutinizing these studies in detail, the special benefits of the corresponding empirical studies of this thesis (study I and study II) become obvious. Most other studies do not cover the whole spectrum of potential task complexity, but present learners mostly with two or with three tasks. Additionally, the underlying dimension task complexity is often not theoretically elaborated; instead often related concepts such as task difficulty are used (Schraw & Roedel, 1994). For example, Rouet (2003) had students solve "specific" and "general" search questions which were supposed to differ in element interactivity and thus in complexity according to cognitive load theory (Sweller, van Merriënboer, & Paas, 2005). However, if these questions are mapped to Bloom's revised Taxonomy (Anderson et al., 2001) they can be classified in the same category (*remember*) indicating similar complexity of underlying cognitive operations. To conclude, the results of these studies investigating students' calibration to task complexity (study I and study II) are consistent with those of other empirical studies; but these studies have the additional benefits of using tasks from the whole spectrum of potential complexities that are also based on a sound theoretical model (Bloom's revised Taxonomy).

The third issue – already discussed in chapter 3.5.3 – concerns *text complexity*: The results of study III are consistent with other empirical studies indicating that learners process differently complex texts differently (Maki, Foley, Kajer, Thompson, & Willert, 1990; McNamara, Kintsch, Songer, & Kintsch, 1996; Schommer & Surber, 1986; Weaver & Bryant, 1995). However, scrutinizing these studies in detail, the special benefits of the empirical study of this thesis become obvious. None of the other empirical studies investigated students' concurrent text processing strategies directly; instead they were inferred from students' learning outcome and the calibration of their comprehension monitoring. Furthermore, a multitude of different operationalisations of text complexity was employed: for example, intact text versus a text with deleted letters (Maki, Foley, Kajer, Thompson, & Willert, 1990), texts varying in their Flesch scores (Weaver & Bryant, 1995) or texts varying in coherence (McNamara, Kintsch, Songer, & Kintsch, 1996). To conclude, the results of the study investigating students' calibration to text complexity (study III) are consistent with those of other empirical studies; but this study has the additional benefits of having

text complexity tested empirically a priori (chapter 3.2.8) and of not only considering learning outcome measures but also the concurrent processes of text comprehension.

### *Conclusion*

The first conclusion concerns research methodology: These empirical studies were the first implementation of the suggested extensions of the traditional construct “calibration” and the first transfer of this methodology to this specific new application context. Considering this, the results can be considered especially promising. Consistently, significant calibration was diagnosed with regard to all employed measures. Especially the indices of *relative calibration* can be easily compared with those from other empirical studies using the *traditional calibration paradigm*. One of the highest indices of relative calibration was found by Nelson and Dunlosky (1991) for delayed JOLs (judgments-of-learning):  $G = .90$ . Under most other conditions indices of relative calibration are lower: for example,  $G = .38$  for immediate JOLs (Nelson & Dunlosky, 1991),  $G = .07$  or  $G = .04$  for predictions of text comprehension (Glenberg & Epstein, 1985) or  $G = .31$  or  $G = .46$  for postdictions of text comprehension (Pressley & Ghatala, 1988). Even postdictions for well-defined tasks such as analogies ( $G = .70$  and  $G = .78$ ) or opposites ( $G = .71$  and  $G = .70$ ) are far from perfect (Pressley & Ghatala, 1988). To conclude, the indices of calibration detected in this series of empirical studies would be considered substantial even if compared with these results of the traditional calibration paradigm. This is one indicator that the transfer of methodology was successful. This research methodology is a valuable enrichment to the investigation of learners’ adaptation to external conditions. But not only the measures discussed so far, but also the visualizations derived from the traditional calibration paradigm can be considered important for the research questions of this thesis: Calibration graphs proved especially beneficial to visualize complex patterns of interactions between external and internal conditions.

As most important conclusion to this section, all results consistently indicate that students in fact do monitor the manipulated external conditions and adapt their whole self-regulated learning process accordingly. Although this may seem trivial, other hypothetical scenarios are conceivable: Students might have had one trait-like approach to learning and thus might not have shown any adaptation to external conditions at all (thus, refuting a central assumption of the COPEs-model). Or they might have adapted their learning to task complexity but not to text complexity (probably because they considered this external condition irrelevant). These empirical results indicate that the selected external conditions were in fact relevant for hypertext learning. Consequently, this is one of the key results of this thesis: Students monitor the external conditions *task complexity* and *text complexity* and calibrate their whole self-regulated learning process accordingly.



### 4.3 Which are the Most Important External Conditions?

The COPES-model (Winne & Hadwin, 1998; chapter 2.1) offers no detailed predictions with regard to this question. Furthermore, an empirical answer is not straightforward: Results of study I and study II indicate that students strongly adapt their learning process to *task complexity*. Results of study III indicate that students strongly adapt their learning to *text complexity*. And the empirical studies in this thesis were not explicitly devised to test external conditions against each other. Therefore, only indirect empirical data can be used to discuss this question. Results tentatively indicate that students seem to consider “hard” constraints such as the externally given time more important than task demands (*task complexity*) which in turn seem to be more influential than demands of the learning material (*text complexity*). Subsequently, arguments for this tentative conclusion will be detailed.

#### *Why “Hard” Constraints might be Most Important*

Consider some *logical arguments*: “Hard” constraints like deadlines should take precedence over everything else in the real life. For example, if students had to submit a term paper before a certain deadline, even the best paper would not receive good grades if handed in too late; it would not be considered at all. The same is true for other “hard” constraints such as permitted materials (during an exam one reference book may be allowed, but others not) or explicit formatting guidelines (a term paper has to be single-spaced instead of double-spaced). These “hard” constraints usually leave *no latitude* for interpretation; instead they represent “either-or” conditions which have to be fulfilled. Thus, it is paramount that students consider such conditions most important!

Consider some *empirical indicators*: During the retrospective stimulated recall interviews in study II, students often reported that they paid attention to the “hard” constraints of the setting: Students knew that the whole study was limited to 2 hours, with about 1 hour of solving tasks with the hypertext on “genetic fingerprinting”. Thus, students were very aware of time and some of them explicitly reported that they reduced their efforts to comply with this constraint, especially for more complex tasks. In total, 66 % of the students in study II mentioned such awareness of time. This potential problem is also corroborated by a comparison between students’ self-reports on planned time (study I) and their actual time spent on tasks (study II). Students estimated that they would need approximately 8 minutes to solve a *remember* task ( $M = 7:70$ ,  $SD = 13:12$ ), approximately 10 minutes to solve an *understand* task ( $M = 9:81$ ,  $SD = 11:78$ ) and approximately 38 minutes to solve an *evaluate* task ( $M = 38:22$ ,  $SD = 61:38$ ). In comparison, students’ *enactments* indicate that they on average spent 6 minutes on *remember* tasks (first *remember* task:  $M = 5:08$ ,  $SD = 4:57$ ; second *remember* task:  $M = 4:50$ ,  $SD = 2:14$ ; last *remember* task:  $M = 8:58$ ,  $SD = 6:41$ ), 10 minutes on the *understand* task ( $SD = 5:01$ ), and 24 minutes on the *evaluate* task ( $SD = 11:38$ ). Note that different samples were used in these studies as well as different tasks. Still, it can be con-

cluded, that students' estimates and their enactments are fairly congruent for *remember* and *understand* tasks, but fairly incongruent for the complex *evaluate* task. Students did not use as much time for this complex task as they planned to use, probably because of the perceived time pressure during task solution in study II. Study III did not yield such systematical qualitative results. However, a small sample of students was interviewed and anecdotal evidence indicates that they were aware of time constraints and adapted their learning accordingly (MJMM30: "Because of the time I was not able to read 'biological background' information or other 'appendices'."). This finding is supported by other empirical studies: For example, Thiede and Dunlosky (1999) showed that when college students were put under conditions of time pressure, they actually spent less time reviewing more difficult information and instead focused on easier content (also Veenman & Beishuizen, 2004). Consequently, these findings tentatively indicate that students might have paid more attention to the "hard" constraint time than to *task complexity* or *text complexity*.

#### *Why Task Complexity might be More Important than Text Complexity*

First, consider some *logical arguments*: Task complexity should be considered more important than text complexity because only few tasks *allow for* adaptation to textual demands. Consider an example: Students have a time-constraint simple *remember* task. Students who adapt their learning process to the presented learning material (spend more time on complex material) would not be able to fulfill the task requirements in time because of these "distractions". The same is true for less time-constraint tasks such as writing a term paper: In this case it is still beneficial to employ a selective reading strategy to focus on the topic of the paper and not be "seduced" by the reading material. Only few tasks call for adaptation to textual demands: If preparing for an exam with a *predefined set of literature*, adaptation to textual demands would be beneficial. For example, the exam could contain questions about specific facts or questions that require deeper understanding. The latter kind of task involves building a full situation model about the presented text material, thus more time and effort should be invested in texts dealing with more complex content. Thus, logically the degree to which adaptation to text demands (*text complexity*) is beneficial or detrimental strongly depends on task demands (*task complexity*).

Consider some *empirical indicators*: In *study III* of this thesis the task was explicitly chosen to allow for students' adaptation to *text complexity*. The results consistently indicate that students in fact do calibrate to text complexity on such a task (students' processed fewer nodes on deeper hierarchical levels (PPN = *percentage of processed nodes*) but processed each of these nodes longer (APDN = *average processing duration of nodes*). Although not an explicit research question of *study II* (and therefore *only* reported here), this study allows for a direct comparison of students' adaptation to task and text complexity. On the one hand, students had to solve differently complex tasks (*task complexity*), and on the other hand they used the hierarchical hypertext to solve these tasks (with *text complexity* indicated by the hierarchical

level). Thus, if text complexity was as important as task complexity, signs of students' adaptation to text complexity should be found consistently for all tasks in this study. In order to systematically compare these results to those of study III, students' *percentage of processed nodes* (PPN) and their *average processing duration per node* (APDN) was computed across all learning tasks and for the five learning tasks separately. For PPN, the results are mostly consistent with those from study III: Students processed a higher percentage of nodes from top-levels than from deeper levels. Note though, that this finding may be an artifact of students' navigation strategies because they most often used hierarchical navigation (e.g., "parent") and thus had to traverse top-level nodes to access deeper level nodes. The results with regard to APDN, however, are not consistent with study III: Only for the second *remember* task focusing on content from a complex "level 3" node students' APDN was higher for deeper nodes than for top-level nodes. For most other tasks moderately complex "level 2" nodes were processed longest (across all five tasks; for the complex *evaluate* task and for the *understand* task). Thus, these results tentatively indicate that students do not adapt their learning to text complexity for all tasks alike. Instead they primarily adapt to task demands which is also indicated by students' navigational patterns: While students in study II accessed the hypertext very selectively, strongly focusing on the specific task (indicator of their calibration to task complexity), students in study III tried to get a comprehensive overview of the whole hypertext (also an indicator of their calibration to task demands as that task required to get an overview). Consequently, empirical results tentatively indicate that these external conditions might interact in a way that students only calibrate to *text complexity* if the task (see *task complexity*) allows or demands such adaptation.

### *Conclusion*

Considering these tentative empirical indicators as well as the proposed logical arguments, the COPES-model could be adapted to incorporate these issues: "Hard" constraints should be explicitly considered more important than "soft" constraints that allow for some latitude with regard to interpretation. Furthermore, task demands such as *task complexity* should be considered more important than demands that are linked to the learning material such as *text complexity*. Note though, that many potentially also influential *external conditions* were not explicitly considered within this series of empirical studies. Thus, this proposed order of *external conditions* should be considered work in progress.

## **4.4 Do Internal Conditions Impact these Adaptation Processes?**

Because the COPES-model (Winne & Hadwin, 1998; chapter 2.1) offered no detailed predictions with regard to this question, more detailed predictions were derived from theoretical models and empirical findings for the internal conditions *prior domain knowledge* (chapter

2.5) and *epistemological beliefs* (chapter 2.6). To investigate the impact of these internal conditions empirically, they were included as covariates or predictor variables in the analyses regarding students' *discrimination* and *calibration* (see previous chapters) to *task complexity* (study I and study II) and to *text complexity* (study III). It was determined if internal conditions had beneficial or detrimental *main* effects and if they *interacted* with the external conditions in all stages of the self-regulated learning process.

To give a preview of the most important conclusion for this chapter: All results consistently indicate that these internal conditions (*prior domain knowledge* and *epistemological beliefs*) significantly impact students' whole self-regulated learning processes.

#### 4.4.1 Prior Domain Knowledge

In all empirical studies of this thesis prior domain knowledge was consistently investigated with a classical method of the expert paradigm: expert – novice comparison. Advanced students of biology were selectively recruited as discipline experts (Rouet, Favart, Britt, & Perfetti, 1997) and humanities students were recruited as novices (Chi, 2006). These quasi-experimental groups were compared. The results of this comparison of prior domain knowledge groups are shortly summarized in Table 11 and were extensively discussed in the local discussions.

##### *Learning Outcome Main Effects*

Prior domain knowledge had a positive *main* effect on the *learning outcome* (study II and study III): Biology students consistently excelled. This effect was relatively independent from the operationalisation of learning outcome, for example it was detected for aggregate sum scores as well as for more qualitative sub-scores (e.g., quality of evaluations, study II). Furthermore, this effect is consistent with other empirical results (Ford & Chen, 2000; Lawless, Brown, Mills, & Mayall, 2003; Rouet, 2003; Rouet, Favart, Britt, & Perfetti, 1997) as well as with the predictions of the COPES-model (chapter 2.5.3).

##### *Learning Process Main Effects*

Additionally, some *main* effects with regard to *learning process* variables were detected: Spending less time on single nodes, using less planning (study II) and considering nodes more comprehensible (study III) might either imply that biology students were more precise in their information searches and achieved more comprehension or imply more superficial processing and less critical attitudes. Empirical results tentatively indicate that biology students' strategies were in fact beneficial for the learning outcome (even though they might seem superficial). These effects are consistent with other empirical studies insofar as more prior domain knowledge was associated with more beneficial learning strategies (Lind &

Sandmann, 2003) as well as with the predictions of the COPES-model insofar as biology students enacted less overt metacognitions (chapter 2.5.3).

Table 11: Summary of prior domain knowledge effects (list on the left, verbal summary on the right)

<b>Study I: Calibration to task complexity in the preparatory stages</b>	
<ul style="list-style-type: none"> <li>- Interaction <i>deep processing</i></li> <li>- Calibration <i>estimated concepts</i></li> </ul>	Prior domain knowledge had minimal impact; it enabled biology students to give more fine-grained judgments ( <i>deep processing</i> ) and be better calibrated.
<b>Study II: Calibration to task complexity in the enactment stages</b>	
<ul style="list-style-type: none"> <li>- Main / interaction <i>time for task completion</i> (TTC)</li> <li>- Main / interaction <i>planning</i> (PL)</li> <li>- Worse calibration <i>number of accessed nodes</i> (NAN)</li> <li>- Interaction <i>task difficulty</i> (learning outcome)</li> <li>- Main <i>evaluations, sum-scores</i> (learning outcome)</li> </ul>	Prior domain knowledge had significant impact; it led to apparently more superficial learning processes, especially for more complex tasks (less TTC, less PL, worse calibration on NAN), but also to superior learning outcome.
<b>Study III: Calibration to text complexity in the enactment stages</b>	
<ul style="list-style-type: none"> <li>- Main / interaction <i>comprehensibility ratings</i> (ACR)</li> <li>- Main <i>learning outcome</i></li> </ul>	Prior domain knowledge had minimal impact; it led to more comprehensible node evaluation (ACR), and to superior learning outcome, but was not systematically related to students' learning strategies.

#### *Learning Process Interactions*

Most *learning process* effects were *interactions*. Contrary to predictions, not all of these interactions indicate that prior knowledge was most influential for complex tasks: For example, in study I biology students' judgments differed most from those of humanities students for mid-complex tasks. Other interactions however are consistent with predictions: In study II, the most valid interaction effect (the other interactions might be artifacts, see local discussion) demonstrates that biology students enacted less planning, especially for complex tasks. And the interaction detected in study III also indicates the largest difference between the prior domain knowledge groups for the most complex texts. To conclude, all valid interaction effects in the *enactment* stage are consistent with predictions based on previous empirical studies (MacDonalds & Stevenson, 1998; Salmerón, Kintsch, & Canas, 2006). The effect in the *preparatory* stage is not consistent with these findings but might be interpreted as partly consistent with the predictions of the COPES-model insofar as experts might have been more *accurate* in diagnosing task demands (chapter 2.5.3).

#### *Learning Process Calibration Effects*

Conceptually related to these interactions also some effects on students' *calibration to external conditions* were detected: In study I biology students' judgments were better calibrated to task complexity than humanities students' judgments. This is consistent with the detected interaction effect (see above) because this also indicates that biology students give more

fine-grained and potentially more accurate judgments. However, in study II biology students demonstrated worse calibration to task complexity than humanities students with regard to their content coverage (NAN). As explained in the corresponding local discussion this might be due to the fact that humanities students needed to access more nodes in order to answer the tasks and thus the probability of good calibration might have been higher. To conclude, the most valid calibration effect detected in the *preparatory* stage of learning is consistent with empirical research demonstrating that experts are more accurate at judging task demands (Chi, Glaser, & Rees, 1982) and is also consistent with the predictions of the COPES-model (see above). However, the calibration effect detected in the *enactment* stage points in the opposite direction.

### *Conclusion*

The main pattern of results makes sense: Prior domain knowledge might have elicited only minimal effects with regard to students' judgments (study I) and with regard to their navigation strategies (study III) because these aspects of the self-regulated learning process might primarily require students' domain-general metacognitive knowledge about strategies. Prior domain knowledge thus might have been only relevant for giving more differentiated judgments (study I). On the other hand, dealing with very specific tasks (study II) or giving judgments for specific texts (study III) might be more dependent on domain-specific knowledge: Prior domain knowledge apparently helped students to comprehend the texts' content (study III), disposed them of the need to engage in extensive overt metacognitions (study II), accelerated their task solution (study II), and facilitated correct answers (learning outcome; study II and study III), especially for more complex tasks.

These effects are consistent with the general assumptions of the COPES-model (Winne & Hadwin, 1998) as prior domain knowledge had significant impact on all COPES-facets of learning: on students' *operations* (e.g., study I: deep processing), on their *products* of learning (e.g., study II: learning outcome), on their *evaluations* (e.g., study III: comprehensibility judgments), and on their *standards* (e.g., study I: number of estimated concepts).

Additionally, these results can be used to specify the effects of prior domain knowledge in all stages of self-regulated learning: First, prior domain knowledge had different impact in different stages of learning: In the *preparatory* stage, prior domain knowledge elicited only minimal interaction effects indicating more fine-grained discrimination and better calibration. In the *enactment* stage, prior domain knowledge elicited more effects: beneficial learning strategies (main), more importance of prior domain knowledge for complex tasks and texts (interaction), and worse calibration. Therefore, the COPES-model could be adapted to also predict more significant effects of prior domain knowledge in the enactment stage in comparison to the preparatory stages. Second, prior domain knowledge interacted differently with different external conditions: In the interaction with *task complexity* prior domain knowledge was very influential for students' learning strategies. In the interaction with

*text complexity* prior domain knowledge was irrelevant for students' learning strategies but only impacted their node evaluations. Therefore, the COPES-model could be adapted to also predict more significant effects of prior domain knowledge for tasks of different complexity than for texts of different complexity.

#### 4.4.2 Epistemological Beliefs

Epistemological beliefs were not operationalised consistently in the empirical studies of this thesis. An experimental manipulation of epistemological beliefs was attempted to elicit more "sophisticated" beliefs (*epistemological sensitization*). This was successful in study II. Therefore in this study two experimental groups were compared: students who received a neutral introduction versus those who read an epistemological introduction. The first prototype of this manipulation failed (study I) and no such manipulation was attempted in study III. In these studies students' epistemological beliefs were captured by different types of instruments: domain-general instruments using denotative statements (study I: EBI; study III: WKI) as well as a domain-specific instrument using connotative adjective pairs (study I and study III: CAEB). Meaningful scales were obtained by exploratory factor analyses (e.g., *CAEB-variability*). The effects of these indicators of epistemological beliefs are shortly summarized in Table 12 and were extensively discussed in the local discussions.

##### *Learning Outcome Main Effects*

Epistemological beliefs had inconsistent *main* effects on the *learning outcome* (study II and study III) and these effects seem to be highly task specific: If learning outcome was operationalised most consistently with the corresponding task (study II: quality of argumentation for a complex *evaluate* task; study III: multiple choice questions about main concepts for the task to "learn as much as possible") more "sophisticated" beliefs were consistently beneficial for the learning outcome. However, if more superficial measures were applied (study II: correctness), "sophisticated" beliefs were detrimental for learning outcome. The beneficial effects of "sophisticated" beliefs on the learning outcome are consistent with other empirical results (Schommer, 1990; Schommer, Crouse, & Rhodes, 1992), especially those demonstrating better argumentation (Kardash & Scholes, 1996; Mason & Boscolo, 2004; Mason & Scirica, 2006), and with the predictions of the COPES-model (chapter 2.6.3). The detrimental effects of "sophisticated" beliefs, however, are not consistent with other empirical results; at most some studies finding no relations between epistemological beliefs and learning outcome could be cited (Mason & Boscolo, 2004).

### Learning Process Main Effects

Additionally, a high number of *main* effects with regard to *learning process* variables were detected: More “sophisticated” beliefs were associated with judging all indicators of deep processing more important across all tasks (study I), with employing more metacognitive planning processes across all tasks (study II), with evaluating nodes more critically and with a selective reading strategy for all texts (shorter processing of nodes, processing of more nodes; study III). All effects are consistent with other empirical studies because they indicate superior learning strategies (planned and enacted) for students with “sophisticated” beliefs (Bartholomé, Stahl, Pieschl, & Bromme, 2006; Kardash & Howell, 2000) as well as with the predictions of the COPES-model (chapter 2.6.3)

Table 12: Summary of epistemological beliefs effects (list on the left, verbal summary on the right)

<b>Study I: Calibration to task complexity in the preparatory stages</b>	
<i>CAEB-variability:</i> <ul style="list-style-type: none"> <li>- Main <i>deep processing</i></li> <li>- Main / interaction <i>dealing with multiple information</i></li> <li>- Main <i>classification according to the Bloom-Categories</i></li> </ul> <i>EBI-definitude:</i> <ul style="list-style-type: none"> <li>- Calibration <i>number of concepts</i></li> </ul>	Domain-dependent “sophisticated” epistemological beliefs ( <i>CAEB-variability</i> ) had a significant impact; they led to judging indicators of deep processing more important. Domain-general “sophisticated” epistemological beliefs ( <i>EBI-definitude</i> ) were less influential; they were associated with less calibration.
<b>Study II: Calibration to task complexity in the enactment stages</b>	
<i>Epistemological sensitization:</i> <ul style="list-style-type: none"> <li>- Interaction <i>time for task completion</i> (TTC)</li> <li>- Calibration <i>time for task completion</i> (TTC)</li> <li>- Main / interaction <i>planning</i> (PL)</li> <li>- Main <i>number of words</i> (learning outcome)</li> <li>- Main <i>argumentation, correctness</i> (learning outcome)</li> <li>- Main <i>task difficulty</i> (learning outcome)</li> </ul>	The epistemological introduction that induced more “sophisticated” epistemological beliefs elicited more elaborated learning processes (better calibration TTC, more PL, especially for more complex tasks) and was beneficial for the number of written words and the quality of argumentation, but was detrimental for giving correct answers.
<b>Study III: Calibration to text complexity in the enactment stages</b>	
<i>WKI-simplicity</i> <ul style="list-style-type: none"> <li>- Main / interaction <i>average processing duration of nodes</i> (APDN)</li> <li>- Main <i>percentage of processed nodes</i> (PPN)</li> <li>- Main <i>average comprehensibility ratings</i> (ACR)</li> <li>- Main <i>total test score</i> (learning outcome)</li> </ul> <i>WKI-certainty</i> <ul style="list-style-type: none"> <li>- Interaction <i>percentage of processed nodes</i> (PPN)</li> <li>- Calibration <i>percentage of processed nodes</i> (PPN)</li> </ul>	Most effects were elicited by domain-general scales (WKI). “Sophisticated” beliefs on <i>WKI-simplicity</i> were associated with a selective reading strategy (short APDN, high PPN), judging nodes less comprehensible (ACR) and better learning outcome. “Sophisticated” beliefs on <i>WKI-certainty</i> were associated with accessing more nodes on deeper level (PPN) and with less calibration (PPN).



### *Learning Process Interactions*

Additionally, a high number of *interactions* with regard to *learning process* variables were detected: Contrary to predictions, not all of these interactions indicate that epistemological beliefs were most influential for complex tasks. For example, in study I “sophisticated” students’ judgments did not differ from “naïve” students’ judgments for the most complex tasks while they differed for simpler tasks. Thus, a reverse pattern was detected. All other interactions, however, are consistent with the predictions: In study II, epistemological beliefs especially impacted the time spent on and the planning processes enacted for the most complex *evaluate* task. In study III, epistemological beliefs especially impacted the time spent on single nodes and the percentage of processed nodes for the most complex hyper-text level (“level 3”). Thus, all interactions in the *enactment* stage are consistent with predictions based on theoretical considerations and the COPES-model (chapter 2.6.3) while the interaction detected in the *preparatory* stage is inconsistent with these assumptions.

### *Learning Process Calibration Effects*

Conceptually related to these interactions also some effects on students’ *calibration to external conditions* were detected: In study I and III more “sophisticated” epistemological beliefs were associated with less calibration, while in study II the epistemological introduction elicited superior calibration. As indicated in the local discussions of these studies, the effect in study I might be due to measurement problems of the associated epistemological beliefs scale (EBI) and the effect in study III might be due to a ceiling effect for the most “sophisticated” students. Therefore, these negative effects should not be overinterpreted. Nonetheless, it should be noted, that only the effect for students’ calibration to *task complexity in the enactment stage* is consistent with other empirical results (with the methodology of the traditional calibration paradigm: Schommer, 1990; Schommer, Crouse, & Rhodes, 1992; with the transferred methodology: Stahl, Pieschl, & Bromme, 2006) and with the predictions of the COPES-model (chapter 2.6.3). Effects in the *preparatory* stage as well as with regard to calibration to *text complexity* point in the opposite direction.

### *Issues and Limitations*

One important issue concerns the measurement of epistemological beliefs. Due to the controversially discussed questions within this relatively young research tradition (e.g., concerning the dimensionality of epistemological beliefs; chapter 2.6.1) the measurement of epistemological beliefs is not simple and straightforward. In an attempt to balance these problems different types of instruments to capture epistemological beliefs were utilized. However, stable factor structures were only found for one instrument: Factor analyses of the CAEB yielded a consistent two factor solution with the scales *CAEB-texture* and *CAEB-variability* (Stahl & Bromme, in press). All other instruments were less reliable: The EBI (Jacobson & Jehng, 1999) was administered in study I to capture students’ domain-general

epistemological beliefs. The original factors were reasonable well-replicated in an exploratory study (Stahl, Pieschl, & Bromme, 2006). However, no meaningful factor solution could be detected; in study I only one EBI scale could be retained: *EBI-definitude*. Due to these problems the EBI was not re-administered; instead another domain-general instrument was tested. However, the original factors of the WKI (Wood & Kardash, 2002) could also not be replicated (study III). Therefore, the measurement problems that were anticipated based on a critical review of this field also proved true empirically: No perfect instrument exists yet. Still, even though the detected dimensions did not consistently replicate the original ones, it proved beneficial to use multiple instruments to get deeper insights in the complex interactions between different kinds of epistemological beliefs (denotative versus connotative; domain-general versus domain-specific).

### *Conclusion*

The high number of epistemological beliefs effects clearly indicates that this internal condition is very influential for the whole self-regulated learning process. However, the results are partly inconsistent (with each other and with other empirical results), especially with regard to effects on calibration and learning outcome. Due to the inherent complexity of this construct (chapter 2.6.1) no simple and coherent pattern of results should have been expected. But if only the majority of effects is considered (and little inconsistencies ignored) the overall pattern of results makes sense: “Sophisticated” epistemological beliefs apparently helped students to plan (study I) and enact more elaborate learning strategies (study II) or learning strategies most beneficial for the learning outcome (study III) and they helped students to evaluate text material more critically (study III). Most of these effects were especially pronounced for more complex tasks or texts (study II and study III). More “sophisticated” epistemological beliefs were also associated with better calibration to task complexity (study II). Additionally, “sophisticated” beliefs facilitated written argumentation (study II) and answering a knowledge test (study III).

These effects are consistent with the general assumptions of the COPES-model (Winne & Hadwin, 1998) as epistemological beliefs had significant impact on all COPES-facets of learning: on students *operations* (e.g., study I: deep processing), on their *products* of learning (e.g., study II: learning outcome), on their *evaluations* (e.g., study III: comprehensibility judgments), and on their *standards* (e.g., study I: estimated concepts).

Additionally, these results can be used to specify the effects of epistemological beliefs in all stages of self-regulated learning: First, epistemological beliefs have different impact in different stages of learning: In the *preparatory* stage, epistemological beliefs elicited primarily main effects. In the *enactment* stage, epistemological beliefs elicited main and interaction effects in almost equal number. Therefore, the COPES-model could be adapted to also predict this differential pattern of effects. Second, epistemological beliefs interact differently with different external conditions and these effects might be related to the issue of

domain-specificity: In the interactions with *task complexity domain-dependent* epistemological beliefs (CAEB, study I) and the domain-dependent epistemological sensitization (study II) were most influential. In the interactions with *text complexity domain-general* epistemological beliefs (WKI, study III) were most influential. Probably, dealing with different texts required more domain-general learning strategies while dealing with very specific tasks required more domain-specific strategies. The COPES-model could be adapted to also predict more significant effects of domain-specific epistemological beliefs for task complexity and more significant effects of domain-general epistemological beliefs for text complexity.

#### 4.5 Do Conditions Impact All Stages Alike?

This was an open question because on the one hand the COPES-model (Winne & Hadwin, 1998) theoretically assumes distinct stages of self-regulated learning which would imply significant differences between stages of self-regulated learning. This assumption is shared by other theoretical models of self-regulated learning (Boekaerts, 1992, 1995, 1996; Borkowski & Burke, 1996; Borkowski, Chan, & Muthukrishna, 2001; Pressley & Ghatala, 1990; Pintrich, 2000; Zimmerman, 1989, 2002). However, these theoretical assumptions were rarely tested empirically except for studies about the validity of self-report instruments which clearly indicate that learners do not always do what they say they do (Jamieson-Noel & Winne, 2003). On the other hand the COPES-model proposes similar learning processes for all stages alike (e.g., the same COPES facets) which would imply similar effects in all stages of self-regulated learning. To detangle the potentially different impact of conditions in different stages of learning, the empirical studies within this thesis either exclusively concentrated on the *preparatory* stages (study I) or primarily focused on the *enactment* stages (study II and study III).

This question can be answered empirically by comparing the results of study I with those of study II. Results with regard to students' calibration to the *external condition task complexity* indicate consistent results in both studies: Students plan (study I) and enact (study II) different strategies for tasks of different complexity. Therefore, students do not seem to suffer profoundly from any fundamental deficit (Winne, 2005, 2006): Results of the preparatory stage indicate that students were able to adequately monitor task complexity (no monitoring deficit) and apparently possessed adequate metacognitive knowledge to plan well-suited strategies (no metacognitive knowledge deficit). Additionally, the results of the enactment stage indicate that students also were able to enact adequate strategies (no production deficit) and were motivated enough to do so (no motivation deficit).

Results with regard to the impact of the *internal condition prior domain knowledge* indicate partly inconsistent results (see previous chapter): In the preparatory stage of learning only minimal impact of prior domain knowledge was detected. In the enactment stage on the

other hand prior domain knowledge had a more significant impact on all learning process variables as well as on students' learning outcome. A potential explanation concerns the fact that students might rely more on domain-independent metacognitive skills to plan adequate strategies (preparatory stage) while the execution of these strategies requires more domain-specific skills (enactment stage). Most data sources indicate that this may not be due to a motivational deficit (because humanities students spent more time on tasks and engaged in more "planning"), but rather due to a production deficit (despite these indicators of deep processing humanities students' learning outcome indicated less success).

Results with regard to the impact of the *internal condition epistemological beliefs* also indicate partly inconsistent results (see previous chapter): In the preparatory stages as well as in the enactment stages of self-regulated learning more "sophisticated" epistemological beliefs were consistently associated with more deep processing strategies. However, differences are also noteworthy: "Sophisticated" beliefs were associated with judging deep processing strategies more important across all tasks in the preparatory stages of learning (*main effect*). On the other hand the epistemological introduction that induced more "sophisticated" beliefs led to deeper processing only for complex tasks (*interaction effect*). Two tentative explanations for these differences might be feasible. The first explanation concerns the validity of measurements of epistemological beliefs: In study I epistemological beliefs were captured by scales measuring students' personal beliefs on a continuum from *absolutist* (knowledge is absolutely certain) to *relativist* (all knowledge is uncertain). The epistemological sensitization administered in study II on the other hand was intended to prompt an *evaluativist* view (some pieces of knowledge are better supported by evidence and thus can be considered more certain than others). Consequently, "sophisticated" students with a *relativistic* view might consider all kinds of deep elaboration strategies important for all kinds of tasks while students with an *evaluativistic* view might be more flexible in their self-regulated learning. The second explanation concerns the validity of measurements of learning strategies: While it is easy to say that one *would apply* deep elaboration strategies even to very simple task, the *enactment* of this plan is expensive and affords cognitive as well as motivational resources. Due to these economic considerations students might not have enacted deep elaboration strategies for simple tasks despite their initial plans.

These discussions highlight another special benefit of this series of empirical studies: By detangling the effects of internal conditions in the preparatory stages and in the enactment stages of learning, it was possible to determine their differential effect. Most previous empirical studies on the other hand neglected the preparatory stage. Therefore, the distinct stages of the COPES-model could be retained, but could be emphasized more strongly.

#### 4.6 Is More Flexible and Accurate Adaptation Beneficial for Learning?

The COPES-model (Winne & Hadwin, 1998; chapter 2.1) posits that adapting to external conditions (*calibration*) constitutes good self-regulation and thus should also be beneficial for the *learning outcome*. However, because the COPES-model itself is an idealized prescriptive model rather than a realistic description of everyday self-regulated learning processes, it is an open issue if these assumptions prove true empirically.

##### *Relations between Calibration Indices and Learning Outcome*

The empirical studies concerned with students' enactment stages contribute relevant results (study II and study III): In both studies *calibration indices* for learning process variables were related to students' learning outcome. In study II strong calibration of *time for task completion* (TTC) to *task complexity* was associated with less correctly solved tasks. In study III, no significant relationship between calibration to *text complexity* and indicators of learning outcome could be determined. Therefore, these results clearly indicate *no beneficial effect* of calibration to external conditions. Instead, it remains an open issue if calibration might not be related to learning outcome at all (as in study III) or if it might be detrimental for the learning outcome (as in study II).

##### *Are these Results Limited to Superficial Learning Outcome Measures?*

One explanation for these counterintuitive results might be feasible: The detected effects might be limited to more superficial measures of learning outcome like those employed in these studies (correctness). For deep conceptual indicators of learning, effects might be different. However, only indirect evidence for this hypothesis can be given.

First, in study II not only the overall correctness of students' task solutions was analyzed, but also the total score given for the complex *evaluate* task that also encompassed qualitative aspects such as the quality of argumentation (chapter 3.4.2.4). In order to better estimate the relative importance of calibration for the learning outcome, other potential determinants of the learning outcome were also tested as predictors of learning outcome: students' *internal conditions* (prior domain knowledge, epistemological beliefs) and *learning process variables* that were used as dependent variables in other analyses (TTC, NAN, PL, EN, REV). These variables had different impact on correctness (superficial) and total score (indicating deep learning): The epistemological sensitization, more time per task (TTC) and more planning (PL) were detrimental for *correctness*. On the other hand the epistemological sensitization was beneficial for *argumentation*, and more time per task (TTC) and more enactment (EN) were beneficial for the *overall score*. Thus, these results clearly indicate that the effects on learning outcome are strongly dependent on outcome measurement.

Second, if the calibration indices of study II are correlated with the overall score for the *evaluate* task, with all constituent sub-scores, as well as with the number of written words

for this task as more qualitative indicators of the learning outcome, the following effects can be found: Calibration indices are not correlated with the overall score or the argumentation sub-score. Consistent with the detrimental effects on the overall correctness, calibration indices for time for task completion (TTC) and enactment (EN) are negatively correlated with the sub-score for *correct conclusions* (TTC:  $r = -.53$ ,  $p = .001$ ; EN:  $r = -.32$ ,  $p = .063$ ). However, better calibration with regard to *time for task completion* (TTC) and planning (PL) was also positively associated with more *written words* (TTC:  $r = .51$ ,  $p = .002$ ; PL:  $r = .40$ ,  $p = .016$ ). Thus, these analyses demonstrate a differential impact of calibration on more superficial learning outcome scores and scores indicating deep learning. Note though that these analyses were *not reported previously*, because of different levels of granularity: While calibration was determined across all five tasks, these more qualitative measures of learning outcome could only be determined for the complex *evaluate* task.

To conclude: The empirically detected detrimental effect of calibration (study II) might be specific to more superficial measures of learning outcome. However, the empirically detected lack of effects (study III) was also found for most qualitative variables indicating deep learning (study II). Therefore, most likely there is no consistent systematic relationship between students' calibration to external conditions and their learning outcome. Only the reported results with regard to the number of written words indicate that calibration might be beneficial for some quantitative learning outcome measures. One central question of this thesis ("*To calibrate or not to calibrate?*") can not be answered finally yet.

#### *Do these Results Imply that Calibration Should not be Fostered?*

Due to the lack of consistent relations between *calibration* and *learning outcome* (see above), it is an open issue if it should be tried to foster better calibration to external conditions. Multiple issues need to be considered: First, the presented results so far only indicate students' *relative calibration*. With regard to the *absolute fit* between students' learning processes and external conditions, it can be assumed that students are far from absolutely well-calibrated: For example, results of study I indicate that students consistently underestimated complexity of very complex tasks and results of study II indicate that they might also have employed too superficial strategies. Second, the results of studies II and III tentatively indicate that calibration to *task complexity* may take precedence over calibration to *text complexity* because not all tasks allow for adaptation to text complexity. A first tentative conclusion can be drawn: Even though it would be possible theoretically to foster better calibration to all kinds of external conditions (*task complexity* or *text complexity*) a more differentiated diagnosis of a specific learning situation is called for. Only calibration to the most relevant external conditions should be fostered. In these contexts calibration might also have positive impact on the learning outcome, especially if indicators of deep learning are considered as learning outcome. Thus, in most learning scenarios, it might be desirable to scaffold adequate cali-

bration to *task complexity*, but with regard to *text complexity* no such clear-cut conclusion should be drawn.

#### 4.7 Proposing Specifications of the COPES-Model

As suggested before, the COPES-model (Winne & Hadwin, 1998) was a valuable heuristic for experimental design and for the operationalisation of relevant variables. Additionally, the results of the series of empirical studies are consistent with the core assumptions of the COPES-model (see above). However, the results can also be used to make more precise specifications about impact of *external* and *internal conditions* in specific stages of self-regulated learning. Therefore, an adapted version of the COPES-model is suggested that is specifically tailored to the research questions and empirical results of this series of studies.

##### *The COPES-Model as Valuable Heuristic*

The COPES-model directly inspired experimental design. First, because of the assumption of distinct stages of self-regulated learning, the studies of this thesis either exclusively focused on the *preparatory* stages of learning or on the *enactment* stages. This design enabled the study of effects of external and internal conditions in different stages. Note though that it was not possible to detangle all assumed stages experimentally (e.g., task definition vs goal setting and planning). However, the division into two broad stages is consistent with the planning and enactment stages in most other models of self-regulated learning (Boekaerts, 1992, 1995, 1996; Borkowski & Burke, 1996; Borkowski, Chan, & Muthukrishna, 2000; Pressley & Ghatala, 1990; Pintrich, 2000; Zimmerman, 2002; for reviews see Montalvo & Torres, 2004; Puustinen & Pulkkinen, 2001) as well as with earlier versions of the COPES-model (Winne, 1996, 1997).

Second, the central assumption of *metacognitive* monitoring and controlling processes to help students *adapt* self-regulated learning to relevant conditions served as initial inspiration for the core questions of this thesis. Even though the COPES framework proved too general for the deduction of specific hypothesis and therefore was supplemented with a more detailed methodology derived from the traditional calibration paradigm, it was the initial inspiration for the topic: “*Conditions and processes of metacognitive calibration*”.

Additionally, the COPES-model was utilized as crucial heuristic to select and evaluate relevant dependent variables in all empirical studies: The COPES-questionnaire used in study I encompasses detailed questions about students’ evaluation of the COPES facets *conditions, operations, standards, and evaluations*. By asking these questions, detailed insights into students’ *preparatory* stages of learning (*task definitions, goals and plans*) were gained. The COPES-model also served as a guiding heuristic in the *enactment* stage. The questions asked in the retrospective stimulated recall interview in study II were explicitly based on the

stages of the COPES-model. Additionally, the coding system for students' concurrent thoughts in study II was impacted by the COPES-model. "Planning" processes can be mapped to students' *preparatory* stages of self-regulated learning, "enactment" processes are congruent with the *enactment* stage of self-regulated learning and "revision" pertains to students' *adaptation* processes. All other utilized dependent variables can also be matched to the COPES-model: Hypertext navigation can be mapped to *operations*, and students' comprehensibility judgments (study III) can be mapped to *evaluations*.

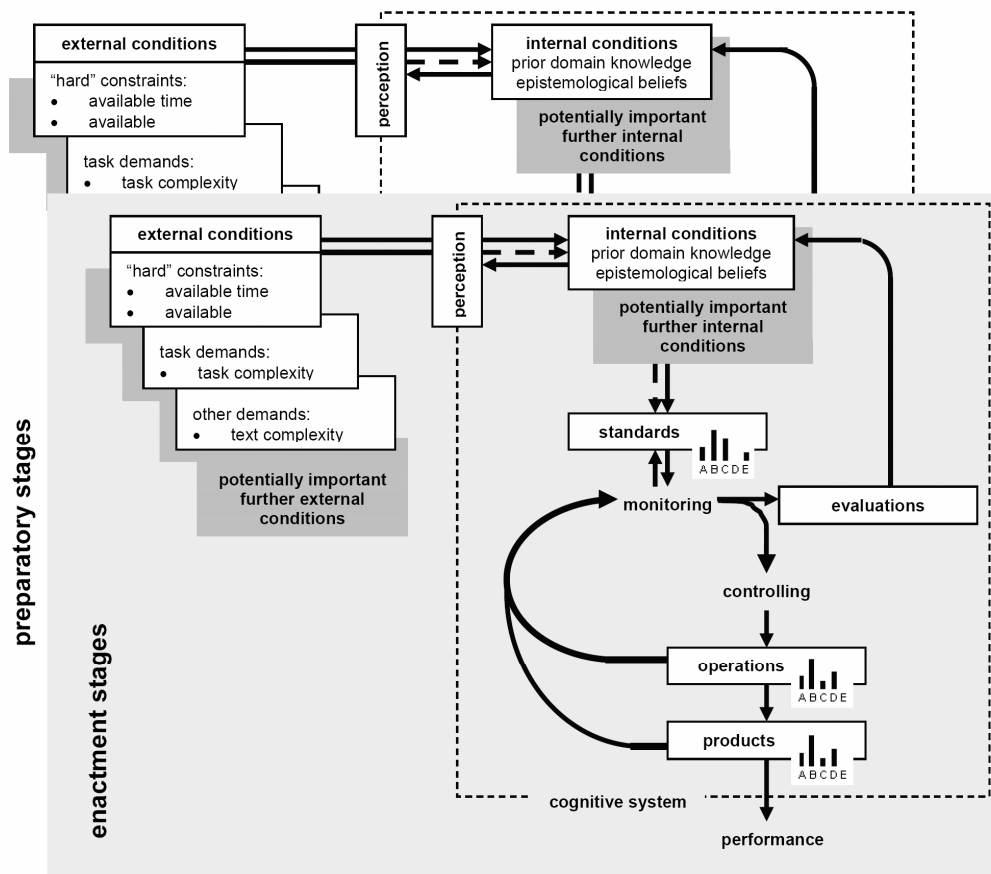


Figure 4.7-1: An adapted version of the COPES-model, specifically tailored to the results of the presented series of empirical studies.

#### *An Adapted and more Specific Version of the COPES-Model*

Because all results of the empirical studies of this thesis were consistent with the COPES-model's core assumptions (see chapter 4.1) the basic structure of the model was retained (COPES facets, metacognitive processes). Two significant changes in comparison with the original model were implemented: First, while the original COPES-model lists all supposedly relevant conditions undifferentiated, the adapted version is specific to the results presented within this thesis and explicitly only lists the investigated conditions (but also ac-



knowledges that other conditions might be equally influential). Second, the investigated conditions elicited different effects in different stages of self-regulated learning; the existence of such stages was further emphasized.

With regard to *external conditions*, two important conditions for hypermedia learning were scrutinized: *task complexity* and *text complexity* (see chapter 4.2). A conceptual analysis and empirical results tentatively indicate that students most strongly adapt their learning to “hard” constraints of the learning situation like available time. Then, they consider task demands like task complexity. Students only systematically adapted their learning to text complexity if tasks allowed for such an adaptation. Thus, the COPES-model can be further specified to explicitly order external conditions according to importance: “hard” constraints > task complexity > text complexity. To visualize this specification (Figure 4.7-1) these external conditions were hierarchically ordered in a stepwise fashion (external condition rectangles in the figure). Note, that this should be considered work in progress as multiple other potentially relevant external conditions were not systematically investigated. To visualize this acknowledgement, gray shades were implemented around these external conditions and labeled “potentially relevant further external conditions”.

With regard to *internal conditions*, two important internal conditions were tested: *prior domain knowledge* and *epistemological beliefs*. Empirical results demonstrate that these internal conditions had differential impact. They elicited partly different effects in different stages of self-regulated learning (preparatory stage vs. enactment stage), with regard to different facets of learning (e.g., operations vs. products), and in interaction with different external conditions (task complexity vs. text complexity) (see chapter 4.4). *Prior domain knowledge* had minimal impact in the preparatory stage (study I) as well as in interaction with text complexity (study III), while it significantly impacted the enacted learning strategies in interaction with task complexity (operations, study II) and the learning outcome (product, study II and study III). *Epistemological beliefs* elicited different patterns of results in different stages: main effects in the preparatory stage (study I) and main and interaction effects in the enactment stage (study II and study III). Additionally, students’ domain-dependent epistemological beliefs were influential for the interaction with task complexity (study I and study II), while their domain-general beliefs were influential for the interaction with text complexity (study III). All these effects can not be visualized in detail in a figure, but some indicators can be incorporated in the adapted model. First, this differentiated pattern calls for a distinction between *stages* of self-regulated learning: Therefore, the different layers for the preparatory and the enactment stage of self-regulated learning indicate that the interaction between all COPES facets (including external and internal conditions) can be of different quality in different stages (Figure 4.7-1). Second, external conditions seem to be perceived through the *filter of learners’ internal conditions*. This is also proposed by the COPES-model, but in this adapted version external and internal conditions were explicitly separated to emphasise their differential impact and to emphasize that external conditions do not im-

pact self-regulated learning directly, but mediated by learners' internal conditions and their perception (Figure 4.7-1). Note, that these assumptions should be considered work in progress as multiple other potentially relevant internal conditions were not systematically investigated. To visualize this acknowledgement, a gray shade was implemented around these internal conditions and labeled "potentially relevant further internal conditions".

Note that empirical results could also be used to further specify the relationship between learning processes and outcome: No systematic relationship between students' adaptation to external conditions and their learning outcome was detected (chapter 4.6). Therefore, the relationship between students' enacted operations and their product and performance – as indicated in the model's visualization – should also be considered tentative. However, this was not explicitly visualized (e.g., by dashed arrows) because other more straightforward operations always demonstrated significant impact on the learning outcome (e.g., selective reading strategy in study III).

#### 4.8 Practical Implications

First, empirical results show differential effects in different stages of self-regulated learning indicating that these stages constitute distinct entities. Additionally, most of the effects detected in the *preparatory* stages of self-regulated learning might transfer to subsequent *enactment* stages. For example, students already calibrated to task complexity in the preparatory stages (study I) as well as in the subsequent enactment stages of self-regulated learning (study II). Contrary to these findings, most empirical studies and educational practices almost exclusively concentrate on the enactment stages of self-regulated learning. This equals giving away an important opportunity: Study I demonstrated significant calibration to task complexity and significant impact of prior domain knowledge and epistemological beliefs even in the preparatory stages. This implies that the preparatory stage of self-regulated learning should receive more attention. If scaffolds or trainings only focus on students' enactments these should fail if students start with inadequate premises about task demands, goals and plans. On the other hand interventions helping students to adequately diagnose task demands in the planning stage and interventions supplying them with conditional knowledge about strategies should be especially helpful. Such an intervention might scaffold adequate task definitions, goals and plans and compensate the detrimental effects of low prior domain knowledge or "naïve" epistemological beliefs.

Second, empirical results regarding students' learning processes in the *enactment* stages mostly indicate that the majority of students tried to give their best (e.g., their concurrent thoughts indicate deep involvement in the subject matter, study II). However, their best might be far from perfect (see above). For example, empirical results from study II and study III consistently demonstrate that accessing "problem" nodes is beneficial for the

learning outcome, probably because such nodes facilitate critical evaluation of the main content of the hypertext. On the other hand empirical results also indicate that especially these nodes are rarely accessed. Note that this finding is analogous to the finding in the help seeking literature that students do not use help functions sufficiently (Aleven, Stahl, Schworm, Fischer, & Wallace, 2003). These issues indicate that besides potential problems to adequately monitor task demands and a potential lack of conditional knowledge about adequate learning strategies in the *preparatory* stage of learning (see above), students' *enactment* stage also might need instructional support.

A third potentially relevant issue in educational practice concerns the *epistemological sensitization*. This instructional intervention elicited more "sophisticated" epistemological beliefs, more deep elaboration learning processes, better calibration, and better argumentation (but less correct answers; study II). However, some open issues with regard to this intervention remain: What are the long-term effects? How exactly does such an intervention interact with students' pre-existing beliefs? What kind of effects would such an intervention elicit outside the laboratory, for example in the classroom? What kinds of comments within the "epistemological" introduction were especially salient and effective? Did the "epistemological" introduction really change students' fundamental and lasting epistemological beliefs or primarily change students' epistemological resources in situ (Hammer & Elby, 2003)? With regard to the last question a third explanation is also feasible: The empirically observed change in beliefs might rather reflect a more contextualized definition of "molecular genetics". Before reading the introduction a student might have thought – correctly – of "molecular genetics" as a wide field that deals with the structure and function of genes at a molecular level with Mendel's laws as typical examples. After reading the introduction this student might have narrowed her notion about what the experimenter might consider relevant for "molecular genetics": now the structure of DNA, the human genome and methods for genetic fingerprinting could be considered typical examples. The semantic space that a student refers to when making her judgments would have shifted. To conclude, on the one hand it is yet unclear how exactly the epistemological sensitization operates and therefore further research into this issue is recommended before implementing this intervention on a larger scale. On the other hand, this intervention proved partially successful empirically and therefore should be considered a potentially powerful way to foster better self-regulation by eliciting more "sophisticated" epistemological beliefs.

#### 4.9 Future Research

Additional research is needed to finally answer the question if calibration to external conditions is beneficial, detrimental, or neutral for learning ("*To calibrate or not to calibrate?*"). This issue could be investigated experimentally. For example, to investigate if calibration *to task*

*complexity* is beneficial, two training interventions could be tested against each other: One intervention could stress the importance of having *one* good strategy for all kinds of tasks (*no calibration*). Students could be trained to run through the same kind of steps for all tasks alike (e.g., first plan what to do, then enact your plan, and last check if the results). The other intervention could stress the importance of matching *different* strategies to the *specific* task affordances (*calibration*). A corresponding training could teach students heuristics how to systematically analyze the task demands and how to determine which kind of strategies would be adequate. Similar experiments could be developed for *text complexity*.

Further additional research could explore how students can be best stimulated to *demonstrate ideal self-regulation* (including good calibration). The results of the empirical studies of this thesis demonstrated that students are fairly good self-regulated learners, but also showed that students learning processes are far from perfect. However, it is less obvious which factors might be responsible: It could be assumed that the perceived “hard” constraints of the setting (e.g., time) did not allow for better self-regulation. Alternatively, student-specific deficits might have prevented better learning processes (monitoring deficit, metacognitive knowledge deficit, production deficit, or motivation deficit; Winne, 2005, 2006; chapter 2.2.1). To determine the contribution of all of these potential reasons, they could be tested against each other, for example in a study focusing on calibration to task complexity: (1) *no time constraint* (thus eliminating this “hard” constraint), (2) *monitoring support* (during task solution students are constantly reminded to monitor), (3) *metacognitive knowledge support* (before task solution students receive instruction detailing the benefits of multiple strategies), (3) *production support* (before task solution students are trained in a number of adequate strategies), or (5) *motivation support* (either by finding intrinsically motivated students or by evoking extrinsic motivation by offering money for good solutions).

#### *How to Ensure Generalizability of Results*

With regard to the *sample*, university students were selected for this series of study because these highly educated students were assumed to possess a certain amount of self-regulatory ability (they might already possess chiseled self-regulated learning strategies and an enhanced awareness of epistemological issues). It was a conscious decision to first investigate the research questions with an almost ideal sample that was most likely to demonstrate metacognitive calibration. However, it might be questionable if the results can be generalized to other populations, especially to those with less formal education. Therefore, it would be interesting to run similar studies with samples from the whole range of potential abilities. To expand the range of *metacognitive knowledge and learning strategies*, samples with less formal education could be scrutinized: For example, high school students might be able to comprehend the topic of “genetic fingerprinting” but might not have as elaborate learning strategies yet. Additionally, future studies could explicitly test for metacognitive knowledge and learning strategies. To expand the range of *prior domain knowledge* a reverse strategy

would have to be employed. The samples used in this series of studies all possess minimal to moderate prior domain knowledge. Even advanced biology students possess only minimal expertise with regard to the specific topic of genetic fingerprinting. Thus, it would be interesting to include real experts of that specific topic in the studies.

Another kind of generalizability pertains to the *domain*. Based on logical arguments it can be assumed that the same kind of effects would have been detected if a different learning domain from “genetic fingerprinting” had been utilized, at least if that domain had similar properties: (1) it is inherently interesting to students, (2) it encompasses certain facts (structure of DNA) as well as controversial issues (certainty of matches), (3) tasks on all levels of complexity can be easily constructed, and (4) an instructional intervention stressing the epistemological nature of the underlying knowledge is credible if the uncertain and changeable nature of knowledge is emphasized. It might be interesting though to explore how students react to domains with different properties. For example, the process of plant identification is very well-structured: Students have to pass through a decision tree with a number of dichotomous decisions which are described by relevant attributes (e.g., option A: plant possesses blossoms vs. option B: plant does not possess blossoms). Furthermore, plant identification is perceived as more traditional than genetic fingerprinting with certain and unchanging knowledge. This perception makes sense as the main book used for plant identification was written about 100 years ago. However, according to experts many inherent controversial issues exist even within this traditional topic of plant identification.

A third kind of generalizability pertains to the *learning material*. In the studies of this thesis students searched for information in the hypertext on “genetic fingerprinting” while most real-life tasks require integration of material from multiple sources that potentially contradict each other (e.g., searching the internet). One of the most important differences between these scenarios concerns the perceived structure of the learning material: The hypertext on “genetic fingerprinting” was perceived as *one unitary document* (AHDD17: “I thought why should they write something that is later on refuted. It is unlikely that they will publish something that they will disprove in the same book”; AHAA22: “I believed the information I read [...] I assumed that the Book is stringent and does not contain contradictory information in different nodes”). This might have led students to engage in too superficial processing and might have led them to overly rely on the presented information instead of critically evaluating the content. Such a strategy would have been inadequate because the hypertext contains noteworthy inconsistencies: For example something might have been introduced as a fact on an introductory level (“every male possesses one allele per Y-chromosomal locus”), but refuted on a more detailed level (“multi-copy loci constitute an exception; males might possess two or four alleles at such Y-chromosomal loci”). However, in real life students will rarely be presented with only one textbook or with a unitary hypertext to learn about a topic (e.g., for an exam). Rather, one of the desired key competencies is the critical *integration of information from multiple sources*. For internet search tasks, for exam-

ple, potential inconsistencies or contradictions are more salient because the different layouts of different websites immediately indicate different authorship. The theory of document representation (Rouet, Favart, Britt, & Perfetti, 1997), for example, suggests how such tasks should be approached: Learners are assumed to consider not only the content of a document, but also information about the source (e.g., author, setting, and form) and the rhetorical goals (e.g., intent and audience). Additionally, different documents should be compared and related to each other by diagnosing consistence, inconsistency, overlap, support, and so on. Thus, it would enhance external validity of the results, if similar effects could be detected with more “open” learning material. For example, it could be expected that the same content of “genetic fingerprinting” would have been perceived differently if presented as internet pages (with different URLs; see Wittwer, Bromme, & Jucks, 2004).

## References

- Aikenhead, G. S., & Ryan, A. G. (1992). The development of a new instrument: Views on Science-Technology and Society (VOSTS). *Science Education*, 76(5), 447-491.
- Ainley, M. D. (1993). Styles of engagement with learning: Multidimensional assessment of their relationship with strategy use and school achievement. *Journal of Educational Psychology*, 85(3), 395-405.
- Airasian, P. W., & Miranda, H. (2002). The role of assessment in the revised taxonomy. *Theory into Practice*, 41(4), 249-254.
- Aleven, V., Stahl, E., Schworm, S., Fischer, F., & Wallace, R. (2003). Help seeking and help design in interactive learning environments. *Review of Educational Research*, 73(3), 277-320.
- Alexander, P. A., & Dochy, F. J. R. C. (1995). Conceptions of knowledge and beliefs: A comparison across varying cultural and educational communities. *American Educational Research Journal*, 32(2), 413-442.
- Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research*, 58(4), 375-404.
- Anderson, G., & Beal, C. R. (1995). Children's recognition of inconsistencies in science texts: Multiple measures of comprehension monitoring. *Applied Cognitive Psychology*, 9, 261-272.
- Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, 98(3), 409-429.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., et al. (2001). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Anderson, L. W., & Sosniak, L. A. (Eds.). (1994). *Bloom's taxonomy. A forty-year retrospective*. Chicago: University of Chicago Press.
- Andrews, R. M., Kubacka, I., Chinery, P. F., Lightowlers, R. N., Turnbull, D. M., & Howell, N. (1999). Reanalysis and revision of the Cambridge reference sequence for mitochondrial DNA. *Nature Genetics*, 23, 147.
- Anslinger, K., & Rolf, B. (2003). *Analyse der mitochondrialen DNA (mtDNA) zur Identifizierung von Knochen oder Leichenteilen* [Analysis of mitochondrial DNA (mtDNA) for identification of bones or parts of dead bodies]. Paper presented at the Münchner Wissenschaftstage, München.
- Artelt, C. (2000). Wie prädiktiv sind retrospektive Selbstberichte über den Gebrauch von Lernstrategien für strategisches Lernen [How predictive are retrospective self-reports about the use of learning strategies for strategic learning]. *Zeitschrift für Pädagogische Psychologie*, 14(2/3), 72-84.
- Athanassiou, N., McNett, J. M., & Harvey, C. (2003). Critical thinking in the management

- classroom: Bloom's taxonomy as a learning tool. *Journal of Management Education*, 27(5), 533-555.
- Ausubel, D. P., & Robinson, F. G. (1969). *School learning: An introduction to educational psychology*. New York: Holt, Rinehart and Winston.
- Ayersman, D. J. (1995). Effects of knowledge representation format and hypermedia instruction on metacognitive accuracy. *Computers in Human Behavior*, 11(3), 533-555.
- Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews*, 4, 829-839.
- Baker, L. (1984a). Spontaneous versus instructed use of multiple standards for evaluating comprehension: effects of age, reading proficiency, and type of standard. *Journal of Experimental Child Psychology*, 38, 289-311.
- Baker, L. (1984b). Children's effective use of multiple standards for evaluating their comprehension. *Journal of Educational Psychology*, 76(4), 588-597.
- Baker, L. (1985a). Differences in the standards used by college students to evaluate their comprehension of expository prose. *Reading Research Quarterly*, 20(3), 297-313.
- Baker, L. (1985b). How do we know when we don't understand? Standards for evaluating text comprehension. In D. L. Forrest-Pressley, G. E. MacKinnon & T. G. Waller (Eds.), *Metacognition, cognition, and human performance* (pp. 155-205). Orlando: Academic Press.
- Ballstaedt, S.-P. (1997). *Wissensvermittlung - Gestaltung von Lehrmaterial* [Teaching - Instructional design]. Weinheim: Beltz Psychologie Verlag Union.
- Bandura, A. (1989). Self-regulation of motivation and action through internal standards and goal systems. In L. A. Pervin (Ed.), *Goal concepts in personality and social psychology* (pp. 19-85). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Bannert, M. (2003). *Assessment of metacognitive skills by means of thinking-aloud instruction and reflection prompts. Does the method affect the learning performance?* Paper presented at the bi-annual conference of the European Association of Research on Learning and Instruction (EARLI), Padua, Italy.
- Barab, S. A., Bowdish, B. E., Young, M. F., & Owen, S. V. (1996). Understanding kiosk navigation: Using log files to capture hypermedia searches. *Instructional Science*, 24, 377-395.
- Baron, D., Braun, J., Erdmann, A., Erdmann, U., Heinze, R., & Lucius, E. R. (Eds.). (2004). *Genetik. Materialien SII. Biologie* [Genetics. Materials for high school biology]. Braunschweig: Schroedel.
- Bartholomé, T., Stahl, E., & Bromme, R. (2004). *Help-seeking in interactive learning environments: Effectiveness of help and learner-related factors in a dyadic setting*. Paper presented at the International Conference of the Learning Sciences: Embracing diversity in the learning sciences, Santa Monica.



- Bartholomé, T., Stahl, E., Pieschl, S., & Bromme, R. (2006). What matters in help-seeking? A study of help effectiveness and learner-related factors. *Computers in Human Behavior*, 22, 113-129.
- Baxter Magolda, M. B. (1992). *Knowing and reasoning in college students: Gender-related patterns in students' intellectual development*. San Francisco: Jossey Bass.
- Baxter Magolda, M. B. (2002). Epistemological reflection: The evolution of epistemological assumptions from age 18 to 30. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about the nature of knowledge and knowing* (pp. 89-102). Mahwah, NJ: Lawrence Erlbaum Associates.
- Beal, C. R. (1996). The role of comprehension monitoring in children's revisions. *Educational Psychology Review*, 8(3), 219-238.
- Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). *Women's ways of knowing: The development of the self, voice, and mind*. New York: Basic Books.
- Bendixen, L. D. (2002). A process model of epistemic belief change. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 191-208). Mahwah, NJ: Lawrence Erlbaum Associates.
- Bendixen, L. D., & Hartley, K. (2003). Successful learning with hypermedia: The role of epistemological beliefs and metacognitive awareness. *Journal of Educational Computing Research*, 28(1), 15-30.
- Bendixen, L. D., & Rule, D. C. (2004). An integrative approach to personal epistemology: a guiding model. *Educational Psychologist*, 39(1), 69-80.
- Benecke, M. (1997). DNA typing in forensic medicine and criminal investigations: A current survey. *Naturwissenschaften*, 84, 181-188.
- Benecke, M. (2001). Genetischer Fingerabdruck [Genetic fingerprinting]. *Enzyklopädie Naturwissenschaft und Technik* (2<sup>nd</sup> ed.), 1-10.
- Bienmiller, A., Shany, M., Inglis, A., & Meichenbaum, D. (1998). Factors influencing children's acquisition and demonstration of self-regulation on academic tasks. In D. H. Schunk & B. J. Zimmerman (Eds.), *Self-regulated learning: From teaching to self-reflective practice* (pp. 203-224). New York: Guilford Press.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. New York: Academic Press.
- Bloom, B. S. (1994). Reflections on the development and use of the taxonomy. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's taxonomy - A forty-year retrospective* (pp. 1 - 9). Chicago: University of Chicago Press.
- Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: Handbook I: cognitive domain*. New York: David McKay.
- Boekaerts, M. (1992). The adaptable learning process: Initiating and maintaining behavioural change. *Applied Psychology: An International Review*, 41(4), 377-397.
- Boekaerts, M. (1995). Self-regulated learning: Bridging the gap between metacognitive and

- metamotivation theories. *Educational Psychologist*, 30(4), 195-200.
- Boekaerts, M. (1996). Personality and the psychology of learning. *European Journal of Personality*, 10(5), 377-404.
- Boekaerts, M. (1999). Metacognitive experiences and motivational state as aspects of self-awareness: Review and discussion. *European Journal of Psychology of Education*, XIV(4), 571-584.
- Boekaerts, M., & Niemivirta, M. (2000). Self-regulated learning: Finding a balance between learning goals and ego-protective goals. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation*. San Diego, CA: Academic Press.
- Borkowski, J. G., & Burke, J. E. (1996). Theories, models, and measurements of executive functioning: An information processing perspective. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive functioning* (pp. 235-261). Baltimore, MD: Paul H Brooks Publishing.
- Borkowski, J. G., Chan, L. K. S., & Muthukrishna, N. (2000). A process-oriented model of metacognition: links between motivation and executive functioning. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 1-41). Lincoln, Nebraska: Buros Institute of Mental Measurements.
- Bortz, J., & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* [Research methods and evaluation for humanities and social sciences] (3rd ed.). Berlin: Springer.
- Boshuizen, H. P. A., & Schmidt, H. G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive Science*, 16, 153-184.
- Bråten, I., & Strømsø, H. I. (2004). Epistemological beliefs and implicit theories of intelligence as predictors of achievement goals. *Contemporary Educational Psychology*, 29, 371-388.
- Bråten, I., & Strømsø, H. I. (2005). The relationship between epistemological beliefs, implicit theories of intelligence, and self-regulated learning among Norwegian post-secondary students. *British Journal of Educational Psychology*, 75, 539-565.
- Bråten, I., & Strømsø, H. I. (2006). Epistemological beliefs, interest, and gender as predictors of internet-based learning activities. *Computers in Human Behavior*, 22(6), 1027-1042.
- Bråten, I., Strømsø, H. I., & Samuelstuen, M. S. (2005). The relationship between internet-specific epistemological beliefs within internet technologies. *Journal of Educational Computing Research*, 33(2), 141-171.
- Brinkmann, B. (1998). Overview of PCR-based systems in identity testing. In P. J. Lincoln & J. Thomson (Eds.), *Methods in molecular biology, Vol. 98: Forensic DNA profiling protocols*. Totowa, NJ: Humana Press Inc.
- Brinkmann, B. (2004). Forensische DNA-Analytik [Forensic DNA analysis]. *Deutsches Ärz-*

- teblatt*, 101(34-35).
- Brinkmann, B., & Wiegand, P. (1997). *DNA-Technologie in der medizinischen Kriminalistik* [DNA technology in medical forensics]. Lübeck: Schmidt-Römhild.
- Brodersen, K., Anslinger, K., & Rolf, B. (2003). *DNA-Analyse und Strafverfahren* [DNA analysis and prosecution]. München: Verlag C. H. Beck.
- Broekkamp, H., van den Bergh, H., van Hout-Wolters, B. H. A. M., & Rijlaardam, G. (2002). Will that be on the test? Perceived task demands and test performance in a classroom context. *European Journal of Psychology of Education*, XVII(1), 75-92.
- Broekkamp, H., & van Hout-Wolters, B. H. A. M. (2006). *Students' adaptation of study strategies when preparing for classroom tests*. Unpublished manuscript, University of Amsterdam, Netherlands, Graduate School of Teaching and Learning.
- Broekkamp, H., van Hout-Wolters, B. H. A. M., Rijlaardam, G., & van den Bergh, H. (2002). Importance in instructional text: teachers' and students' perception of task demands. *Journal of Educational Psychology*, 94(2), 260-271.
- Broekkamp, H., van Hout-Wolters, B. H. A. M., van den Bergh, H., & Rijlaardam, G. (2004). Teachers' task demands, students' test expectations, and actual test content. *British Journal of Educational Psychology*, 74, 205-220.
- Bromme, R. (1992). *Der Lehrer als Experte: Zur Psychologie des professionellen Wissens* [The teacher as expert: Psychology of professional knowledge]. Bern: Huber.
- Bromme, R. (2005). Thinking and knowing about knowledge: A plea for and critical remarks on psychological research programs on epistemological beliefs. In M. Hoffman, J. Lenhard & F. Seeger (Eds.), *Activity and sign - Grounding mathematics education* (pp. 191-201). New York: Springer.
- Bromme, R., & Stahl, E. (2003). *The impact of epistemological beliefs on e-learning: The case of help-seeking*. Paper presented at the Joint workshop of cognition and learning through media-communication for advanced e-learning, Berlin.
- Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-115). London: Lawrence Erlbaum Associates.
- Bruce, R. L. (1981). *Programming for intangibles*. Ithaca, NY: New York State College of Agriculture and Life Sciences.
- Brunstein, A., & Krems, J. F. (2005). Einfluss des Bearbeitungsziels auf die Strategiewahl beim hypertextgestützten Lernen [Impact of the processing goal on strategy choice during learning with hypertexts]. *Zeitschrift für Pädagogische Psychologie*, 19(1/2), 39-48.
- Budowle, B., & Allen, R. C. (1998). Analysis of amplified fragment-length polymorphisms (VNTR / STR loci) for human identity testing. In P. J. Lincoln & J. Thomson (Eds.), *Methods in molecular biology, Vol. 98: Forensic DNA profiling protocols*. Totowa, NJ: Humana Press Inc.
- Budowle, B., Smith, J., Moretti, T., & DiZinno, J. (2000). *DNA typing protocols: molecular biol-*

- ogy and forensic analysis*. Natrick, MA: Eaton Publishing.
- Buehl, M. M., & Alexander, P. A. (2001). Beliefs about academic knowledge. *Educational Psychology Review*, 13(4), 385-418.
- Buehl, M. M., Alexander, P. A., & Murphy, K. P. (2002). Beliefs about schooled knowledge: Domain specific or domain general? *Contemporary Educational Psychology*, 27, 415-449.
- Burson, K. A., Klayman, J., & Larrick, R. P. (2006). Skilled or unskilled, but still unaware of it: How perception of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology*, 90(1), 60-77.
- Butler, D. L., & Cartier, S. C. (2004). Promoting effective task interpretation as an important work habit: A key to successful teaching and learning. *Teachers College Record*, 106(9), 1729-1758.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.
- Butler, J. M. (1998). The use of capillary electrophoresis in genotyping STR loci. In P. J. Lincoln & J. Thomson (Eds.), *Methods in molecular biology, Vol. 98: Forensic DNA profiling protocols*. Totowa, NJ: Humana Press Inc.
- Butler, J. M. (2001). Chapter 1: Overview and history of DNA typing. In J. M. Butler (Ed.), *Forensic DNA typing : Biology & technology behind STR markers*: Academic Press.
- Butler, J. M. (2005). *PowerPoint slides for use with "Forensic DNA Typing", 2nd edition*. Retrieved 11th June, 2007, from <http://www.cstl.nist.gov/biotech/strbase/FDT2e.htm>
- Butler, J. M., Buel, E., Crivellente, F., & McCord, B. R. (2004). Forensic DNA typing by capillary electrophoresis using ABI prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis*, 25, 1397-1412.
- Butler, J. M., Decker, A. E., Vallone, P. M., & Kline, M. C. (2005). Allele frequencies for 27 Y-STR loci with U.S. Caucasian, African American, and Hispanic samples. *Forensic Science International*, 156, 250-260.
- Butler, J. M., Schoske, R., Vallone, P. M., Redman, J. W., & Kline, M. C. (2003). Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American and Hispanic populations. *Journal of Forensic Sciences*, 48(4), 1-4.
- Butler, J. M., Shen, Y., & McCord, B. R. (2003). The development of reduced size STR amplicons as tools for analysis of degraded DNA. *Journal of Forensic Sciences*, 48(5), 1054-1064.
- calibration*. Retrieved 1st June, 2007, from <http://en.wikipedia.org/wiki/Calibration>
- calibrate*. Retrieved 1st June, 2007, from <http://www.merriam-webster.com/dictionary/calibrate>
- calibrate*. Retrieved 1st June, 2007, from <http://wordnet.princeton.edu/perl/webwn?s=calibrate&sub=Search+WordNet&o2=&o0=1&o7=&o5=&o1=1&o6=>

- &o4=&o3=&h=
- Calisir, F., & Gurel, Z. (2003). Influence of text structure and prior knowledge of the learner on reading comprehension, browsing and perceived control. *Computers in Human Behavior*, *19*, 135-145.
- Campbell, D. J. (1988). Task complexity: A review and analysis. *Academy of Management Review*, *37*, 60-82.
- Cano, F. (2005). Epistemological beliefs and approaches to learning: Their change through secondary school and their influence on academic performance. *British Journal of Educational Psychology*, *75*, 203-221.
- Canter, D., Rivers, R., & Storrs, G. (1985). Characterizing user navigation through complex data structures. *Behavior & Information Technology*, *4*(2), 93-102.
- Carey, L., & Mitnik, L. (2002). Trends in DNA forensic analysis. *Electrophoresis*, *23*, 1386-1397.
- Carracedo, A., Bär, W., Lincoln, P. J., Mayr, W., Morling, N., Olaisen, B., et al. (2000). DNA Commission of the International Society for Forensic Genetics: Guidelines for mitochondrial DNA typing. *Forensic Science International*, *110*, 79-85.
- Carroll, M., & Korukina, S. (1999). The effect of text coherence and modality on metamemory judgments. *Memory*, *7*(3), 309-322.
- Carver, C. S., & Scheier, M. (1990). Principles of self-regulation: Action and emotion. In E. T. Higgins & R. M. Sorrentino (Eds.), *Handbook of motivation and cognition: Foundations of social behavior* (Vol. 2). New York, NY: Guilford Press.
- Chan, C. C., Tsui, M. S., Chan, M. Y., & Hong, J. H. (2002). Applying the structure of the observed learning outcomes (SOLO) taxonomy on student's learning outcomes: An empirical study. *Assessment and Evaluation in Higher Education*, *27*(6), 511-527.
- Chandler, M. J., Hallett, D., & Sokol, B. W. (2002). Competing claims about competing knowledge claims. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 145-168). Mahwah, NJ: Lawrence Erlbaum Associates.
- Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning & Memory*, *2*(6), 641-653.
- Charness, N., Tuffiash, M., Krampe, R., Reingold, E., & Vasyukoya, E. (2005). The role of deliberate practice in chess expertise. *Applied Cognitive Psychology*, *19*, 151-165.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55-81.
- Chi, M. T. H. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 21-30). New York: Cambridge University Press.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152.
- Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Stern-

- berg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1, pp. 7-75). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science. *Review of Educational Research, 63*(1), 1-49.
- Clarebout, G., & Elen, J. (2001). The ParlEuNet-project: Problems with the validation of socio-constructivist design principles in ecological settings. *Computers in Human Behavior, 17*(5-6), 463-464.
- Clarebout, G., Elen, J., Luyten, L., & Bamps, H. (2001). Assessing epistemological beliefs: Schommer's questionnaire revisited. *Educational Research and Evaluation, 7*(1), 53-77.
- Clark, R. E., & Feldon, D. F. (2005). Five common but questionable principles of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 97-115). New York: Cambridge University Press.
- Clinchy, B. M. (2002). Revisiting women's ways of knowing. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 63-88). Mahwah, NJ: Lawrence Erlbaum Associates.
- Commander, N. E., & Stanwyck, D. J. (1997). Illusion of knowing in adult readers: Effects of reading skill and passage length. *Contemporary Educational Psychology, 22*(1), 39-52.
- Cordón, L. A., & Day, J. D. (1996). Strategy use on standardized reading comprehension tests. *Journal of Educational Psychology, 88*(2), 288-295.
- Cornacchia, D., & Fitch, I. (2005). Introducing Y-STR DNA testing in court. *Profiles in DNA, 8*(1), 6-8.
- Daghio, M. M., Fattori, G., & Ciardullo, A. V. (2006). Assessment of readability and learning of easy-to-read educational health materials designed and written with the help of citizens by means of two non-alternative methods. *Advances in Health Sciences Education, 11*, 123-132.
- Dahl, T. I., Bals, M., & Turi, A. L. (2005). Are students' beliefs about knowledge and learning associated with their reported use of learning strategies? *British Journal of Educational Psychology, 75*, 257-273.
- Dalgety, J., Coll, R. K., & Jones, A. (2003). Development of chemistry attitudes and experiences questionnaire (CAEQ). *Journal of Research in Science Teaching, 40*(7), 649-668.
- Davidson, J. E., & Sternberg, R. J. (1998). Smart problem solving: How metacognition helps. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 47-68). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- De Corte, E. (1973). *Onderwijsdoelstellingen*. Louvain: Universitaire Pers.
- De Corte, E., Op 't Eynde, P., & Verschaffel, L. (2002). "Knowing what to believe": The relevance of students' mathematical beliefs for mathematics education. In B. K.

- Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 297-320). Mahwah, NJ: Lawrence Erlbaum Associates.
- De Jong, T., & Ferguson-Hessler, M. G. M. (1996). Types and qualities of knowledge. *Educational Psychologist, 31*(2), 105-113.
- De Knijff, P. (2003). Son, give up your gun: Presenting Y-STR results in court. *Profiles in DNA, 3*-5.
- Dempster, F. (1988). The spacing effect: A case study in the failure to apply the results of psychological research. *American Psychologist, 43*(8), 627-634.
- Desoete, A., Roeyers, H., & Buysse, A. (2001). Metacognition and mathematical problem solving in grade 3. *Journal of Learning Disabilities, 34*(5), 435-449.
- Duell, O. K., & Schommer-Aikins, M. (2001). Measures of people's beliefs about knowledge and learning. *Educational Psychology Review, 13*(4), 419-449.
- Dunlosky, J., & Hertzog, C. (2000). Updating knowledge about encoding strategies: A componential analysis of learning about strategy effectiveness from task experience. *Psychology and Aging, 15*(3), 462-474.
- Dunlosky, J., Rawson, K. A., & Hacker, D. J. (2002). Metacomprehension of science texts: Investigating the levels-of-disruption hypothesis. In J. Otero, J. A. León & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 255-279). Mahwah, NJ: Lawrence Erlbaum Associates.
- Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self-paced study. *Acta Psychologica, 98*(1), 37-56.
- Duran, N. D., Bellissens, C., Taylor, R. S., & McNamara, D. S. (2007). *Quantifying text difficulty with automated indices of cohesion and semantics*. In D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of The 29th Annual Conference of the Cognitive Science Society* (pp. 233-238), Nashville, Tennessee, USA.
- Edson, S. M., Ross, J. P., Coble, M. D., Parsons, T. J., & Barritt, S. M. (2004). Naming the dead - confronting the realities of rapid identification of degraded skeletal remains. *Forensic Science Review, 16*(1), 63-90.
- Efklides, A., Samara, A., & Petropoulou, M. (1999). Feeling of difficulty: An aspect of monitoring that influences control. *European Journal of Psychology of Education, XIV*(4), 461-476.
- Efklides, A., & Vauras, M. (1999). Introduction. *European Journal of Psychology of Education, XIV*(4), 455-459.
- Eisenberger, R. (1992). Learned industriousness. *Psychological Review, 99*(2), 248-267.
- Eisenberger, R., Masterson, F. A., & McDermitt, M. (1982). Effects of task variability on generalized effort. *Journal of Educational Psychology, 74*(4), 499-505.
- Elby, A., Frederiksen, J., Schwarz, C., & White, B. (2003). *Epistemological beliefs assessment for physical science (EBAPS)*. Retrieved 11th June, 2007, from

- <http://www2.physics.umd.edu/~elby/EBAPS/home.htm>
- Elby, A., & Hammer, D. (2001). On the substance of a sophisticated epistemology. *Science Education, 85*(5), 554-567.
- Elen, J., & Clarebout, G. (2001). An invasion in the classroom: Influence of an ill-structured innovation on instructional and epistemological beliefs. *Learning Environment Research, 4*, 87-105.
- Enokson, R. (1973). A simplified teacher question classification model. *Education, 94*(1), 27-29.
- Ericsson, K. A. (2005). Recent advances in expertise research: A commentary on the contributions to the special issue. *Applied Cognitive Psychology, 19*, 233-241.
- Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 221-245.
- Ericsson, K. A., Krampe, R., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*(3), 363-406.
- Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215-251.
- Estes, D., Chandler, M., Horvath, K. J., & Backus, D. W. (2003). American and British college students' epistemological beliefs about research on psychological and biological development. *Applied Developmental Psychology, 23*, 625-642.
- Feltovich, P. J., Prietula, M. J., & Ericsson, K. A. (2006). Studies of expertise from psychological perspectives. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 41-67). New York: Cambridge University Press.
- Fitts, P. M. (1964). Perceptual-motor skill learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 243-385). New York: Academic Press.
- Fitzgerald, J. T., White, C. B., & Gruppen, L. D. (2003). A longitudinal study of self-assessment accuracy. *Medical Education, 37*, 645-649.
- Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-236). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Flavell, J. H. (1987). Speculations about the nature and development of metacognition. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 21-29). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Flavell, J. H. (1992). Metacognition and cognitive monitoring. In T. O. Nelson (Ed.), *Metacognition: Core readings* (pp. 3-8). Boston: Allyn and Bacon.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221-233.
- Ford, N., & Chen, S. Y. (2000). Individual differences, hypermedia navigation and learning: An empirical study. *Journal of Educational Multimedia and Hypermedia, 9*(4), 281-311.
- Fourney, R. M. (2002). *Forensic reality and the practical experience of DNA typing update*. Re-



- rieved 20. September, 2004, from <http://www.isrcl.org/Papers/Fourney.pdf>
- Fridell, R. (2001). *DNA fingerprinting: the ultimate identity*. New York: Franklin Watts, Div. Scholastic Inc.
- Furst, E. J. (1994). Bloom's taxonomy: Philosophical and educational issues. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's taxonomy: A forty-year retrospective* (pp. 28 - 41). Chicago: University of Chicago Press.
- Gagné, R. M. (1972). Domains of learning. *Interchange*, 3(1), 1-8.
- Gall, J. (2006). Orienting tasks and their impact on learning and attitudes in the use of hypertext. *Journal of Educational Multimedia and Hypermedia*, 15(1), 5-29.
- Gall, J., & Hannafin, M. J. (1994). A framework for the study of hypertext. *Instructional Science*, 22(3), 207-232.
- Garavalia, L. S., & Gredler, M. E. (2003). An exploratory study of academic goal setting, achievement calibration and self-regulated learning. *Journal of Instructional Psychology*, 29(4), 221-230.
- García-Arista, E., Campanario, J. M., & Otero, J. (1996). Influences of subject matter setting on comprehension monitoring. *European Journal of Psychology of Education*, XI(4), 427-441.
- Garner, R. (1980). Monitoring of understanding: An investigation of good and poor readers' awareness of induced miscomprehension of text. *Journal of Reading Behavior*, XII(1), 55-64.
- Gentner, D., & Markman, A. B. (1994). Structural alignment in comparison: No difference without similarity. *Psychological Science*, 5(3), 152-158.
- Gerlach, V., & Sullivan, A. (1967). *Constructing statements of outcomes*. Inglewood, CA: Southwest Regional Laboratory for Educational Research and Development.
- Gierl, M. J. (1997). Comparing cognitive representations of test developers and students on a mathematics test with Bloom's taxonomy. *The Journal of Educational Research*, 91(1), 26-32.
- Gill, M. G., Ashton, P. T., & Algina, J. (2004). Changing preservice teachers' epistemological beliefs about teaching and learning in mathematics: An intervention study. *Contemporary Educational Psychology*, 29, 164-185.
- Gill, P., Brenner, C., Brinkmann, B., Budowle, B., Carracedo, A., Jobling, M. A., et al. (2001). DNA Commission of the International Society of Forensic Genetics: Recommendations on forensic analysis using Y-chromosome STRs. *International Journal of Legal Medicine*, 114, 305-309.
- Gill, P., Sparkes, R., & Tully, G. (2001). *DNA profiling in forensic sciences*. Retrieved 2. February, 2005, from <http://www.els.net/>
- Glenberg, A. M., & Epstein, W. (1985). Calibration of comprehension. *Journal of Experimental Psychology-Learning Memory and Cognition*, 11(4), 702-718.

- Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory & Cognition*, 15(1), 84-93.
- Glenberg, A. M., Sanocki, T., Epstein, W., & Morris, C. (1987). Enhancing calibration of comprehension. *Journal of Experimental Psychology: General*, 116(2), 119-136.
- Gobet, F. (1996). Expertise und Gedächtnis [Expertise and memory]. In H. Gruber & A. Ziegler (Eds.), *Expertiseforschung: Theoretische und methodische Grundlagen* (pp. 58-79). Opladen: Westdeutscher Verlag.
- Gobet, F. (2005). Chunking models of expertise: Implications for education. *Applied Cognitive Psychology*, 19, 183-204.
- Goodman, L. A., & Kruskal, W. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, 49, 732-769.
- Graham, S., Schwartz, S. S., & MacArthur, C. A. (1993). Knowledge of writing and the composing process, attitude toward writing, and self-efficacy for students with and without learning disabilities. *Journal of Learning Disabilities*, 26(4), 237-249.
- Granello, D. H. (2001). Promoting cognitive complexity in graduate written work: Using Bloom's taxonomy as a pedagogical tool to improve literature reviews. *Counselor Education & Supervision*, 40, 292-307.
- Greene, J. A., & Azevedo, R. (2007). A theoretical review of Winne and Hadwin's model of self-regulated learning: New perspectives and directions. *Review of Educational Research*, 77(3), 334-372.
- Grimm, L. G., & Yarnold, P. (Eds.). (1995). *Reading and understanding multivariate statistics*. Washington: American Psychological Association.
- Groeben, N. (1972). *Die Verständlichkeit von Unterrichtstexten* [Comprehensibility of instructional texts]. Münster: Aschendorff Verlag.
- Gruber, H. (1994). *Expertise. Modelle und empirische Untersuchungen* [Expertise. Models and empirical studies]. Opladen: Westdeutscher Verlag GmbH.
- Gruber, H. (2006). Expertise. In D. E. Rost (Ed.), *Handwörterbuch Pädagogische Psychologie* [Dictionary of educational psychology] (Vol. 3, pp. 164 - 170). Weinheim: Beltz.
- Gruber, H., & Mandl, H. (1996). Expertise und Erfahrung [Expertise and experience]. In H. Gruber & A. Ziegler (Eds.), *Expertiseforschung: Theoretische und methodische Grundlagen* (pp. 18-33). Opladen: Westdeutscher Verlag.
- Gruber, H., & Strube, G. (1989). Zweierlei Experten: Problemisten, Partierspieler und Novizen beim Lösen von Schachproblemen [Two kinds of experts: Problemists', game players' and novices' solutions of chess problems]. *Sprache & Kognition*, 8(2), 72-85.
- Gruber, H., & Ziegler, A. (1996). Expertise als Domäne psychologischer Forschung [Expertise as domain of psychological research]. In H. Gruber & A. Ziegler (Eds.), *Expertiseforschung: Theoretische und methodische Grundlagen* (pp. 7-15). Opladen: Westdeutscher Verlag.
- Gusmao, L., & Carracedo, A. (2003). *Y chromosome-specific STRs*. Retrieved 2. February, 2005,

- from [www.pomega.com/profiles/601/ProfilesInDNA\\_601\\_03.pdf](http://www.pomega.com/profiles/601/ProfilesInDNA_601_03.pdf)
- Guttman, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, 18, 277-296.
- Hacker, D. J. (1997). Comprehension monitoring of written discourse across early-to-middle adolescence. *Reading and Writing: An Interdisciplinary Journal*, 9, 207-240.
- Hacker, D. J. (1998). Self-regulated comprehension during normal reading. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 165-191). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Hadwin, A. F., & Winne, P. H. (2001). CoNoteS2: A software tool for promoting self-regulation. *Educational Research and Evaluation*, 7(2-3), 313-334.
- Hadwin, A. F., Winne, P. H., Stockley, D. B., Nesbit, J. C., & Woszczyna, C. (2001). Context moderates students' self-reports about how they study. *Journal of Educational Psychology*, 93(3), 477-487.
- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). *Multivariate data analysis* (5th ed.). Upper Saddle River, NJ: Prentice-Hall.
- Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. Boston: Allyn & Bacon.
- Halloun, I. (1997). *Views about science and physics achievement. The VASS story*. Paper presented at the International Conference on Undergraduate Physics Education, College Park, MD.
- Hammer, D. (1994). Epistemological beliefs in introductory physics. *Cognition and Instruction*, 12(2), 151-183.
- Hammer, D. (1995). Epistemological considerations in teaching introductory physics. *Science Education*, 79(4), 393-413.
- Hammer, D., & Elby, A. (2002). On the form of a personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 169-190). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hammer, D., & Elby, A. (2003). Tapping epistemological resources for learning physics. *The Journal of the Learning Sciences*, 12(1), 53-90.
- Handt, O., Meyer, S., & von Haeseler, A. (1998). Compilation of human mtDNA control region sequences. *Nucleic Acids Research*, 26(1), 126-129.
- Hannah, L. S., & Michaelis, J. U. (1977). *A comprehensive framework for instructional objectives: A guide to systematic planning and evaluation*. Reading, MA: Addison-Wesley.
- Hare, V. C., & Borchardt, K. M. (1985). Good and poor comprehenders' detection of errors revisited. *The Journal of Educational Research*, 78(4), 237-241.
- Hartley, K., & Bendixen, L. D. (2003). The use of comprehension aids in a hypermedia environment: Investigating the impact of metacognitive awareness and epistemological beliefs. *Journal of Educational Multimedia and Hypermedia*, 12(3), 275-289.

- Hasselhorn, M. (1998). Metakognition [Metacognition]. In D. E. Rost (Ed.), *Handwörterbuch der Pädagogischen Psychologie* (pp. 348-351). Weinheim: Beltz.
- Hauenstein, A. D. (1998). *A conceptual framework for educational objectives: A holistic approach to traditional taxonomies*. Lanham, MD: University Press of America.
- Hidi, S. E. (1995). A reexamination of the role of attention in learning from text. *Educational Psychology Review*, 7(4), 323-350.
- Hofer, B. K. (2000). Dimensionality and disciplinary differences in personal epistemology. *Contemporary Educational Psychology*, 25, 378-405.
- Hofer, B. K. (2001). Personal epistemology research: Implications for learning and teaching. *Educational Psychology Review*, 13(4).
- Hofer, B. K. (2002). Personal epistemology as a psychological and educational construct: An introduction. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 3-14). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hofer, B. K. (2004). Exploring the dimensions of personal epistemology in differing classroom contexts: Student interpretations during the first year of college. *Contemporary Educational Psychology*, 29, 129-163.
- Hofer, B. K. (2004). Epistemological understanding as a metacognitive process: Thinking aloud during online-searching. *Educational Psychologist*, 39(1), 43-55.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67(1), 88-140.
- Hofer, B. K., & Pintrich, P. R. (2002). *Personal epistemology. The psychology of beliefs about knowledge and knowing*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Hohoff, C., & Brinkmann, B. (2003). Trends in der forensischen Molekulargenetik [Trends in forensic molecular genetics]. *Rechtsmedizin*, 13, 183-189.
- Holland, M. M., & Parsons, T. J. (1999). Mitochondrial DNA sequence analysis - validation and use for forensic casework. *Forensic Science Review*, 11(1), 21-50.
- Huitema, B. K. (1980). *The analysis of covariance and alternatives*. New York: Wiley.
- Isenberg, A. R. (2002). Forensic mitochondrial DNA analysis: A different crime-solving tool. *The FBI Law Enforcement Bulletin*.
- Jacobson, M. J., & Jehng, J.-C. (1999). *Epistemological beliefs instrument: Scales and items*. Retrieved June 20, 2007, from [http://mjacobson.net/publications/Epist\\_Beliefs\\_Instrument98.PDF](http://mjacobson.net/publications/Epist_Beliefs_Instrument98.PDF).
- Jacobson, M. J., Maouri, C., Mishra, P., & Kolar, C. (1996). Learning with hypertext learning environments: Theory, design, and research. *Journal of Educational Multimedia and Hypermedia*, 5(3/4), 239-281.

- Jacobson, M. J., & Spiro, R. J. (1995). Hypertext learning environments, cognitive flexibility, and the transfer of complex knowledge: An empirical investigation. *Journal of Educational Computing Research*, 12(4), 301-333.
- Jamieson-Noel, D., & Winne, P. H. (2003). Comparing self-reports to traces of studying behavior as representations of students' studying and achievement. *Zeitschrift für Pädagogische Psychologie*, 17(3/4), 159-171.
- Jehng, J. J., Johnson, S., & Anderson, R. C. (1993). Schooling and students' epistemological beliefs about learning. *Contemporary Education*, 18, 23-35.
- Jobling, M. A., & Gill, P. (2004). Encoded evidence: DNA in forensic analysis. *Nature Reviews*, 5, 739-751.
- Jobling, M. A., Pandya, A., & Tyler-Smith, C. (1997). The Y chromosome in forensic analysis and paternity testing. *International Journal of Legal Medicine*, 110, 118-124.
- Johnson, P. E., Duran, A. S., Hassebrock, F., Moller, J., Prietula, M., J., Feltovich, P. J., et al. (1981). Expertise and error in diagnostic reasoning. *Cognitive Science*, 5, 235-283.
- Jucks, R. (2001). *Was verstehen Laien? Zur Verständlichkeit von Fachtexten aus der Sicht von Computer-Experten* [What do novices comprehend? The comprehensibility of scientific texts from the perspective of computer experts]. Münster: Waxmann.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23-31.
- Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design. *Human Factors*, 40, 1-17.
- Kalyuga, S., Chandler, P., & Sweller, J. (2001). When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, 93, 579-588.
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative / aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, 74(4), 657-690.
- Kardash, C. M., & Howell, K. L. (2000). Effects of epistemological beliefs and topic-specific beliefs on undergraduates' cognitive and strategic processing of dual-positional text. *Journal of Educational Psychology*, 92(3), 524-535.
- Kardash, C. M., & Scholes, R. J. (1996). Effects of preexisting beliefs, epistemological beliefs, and need for cognition on interpretation of controversial issues. *Journal of Educational Psychology*, 88(2), 260-271.
- Keller, C., & Hülsmann, U. (2003). *Der genetische Fingerabdruck. Die DNA-Analyse in der polizeilichen Praxis* [The genetic fingerprint. DNA analysis in police practice] (Vol. 2. revised Ed.). Stuttgart: Richard Boorberg Verlag.
- Kienhues, D., Bromme, R., & Stahl, E. (in press). Changing epistemological beliefs: The surprising impact of a short-term intervention. *British Journal of Educational Psychology*.
- Kim, Y. C. (2003). The effects of strategic processing in reading upon inference generation

- and comprehension. *Korean Journal of Thinking & Problem Solving*, 13(1), 31-48.
- King, P. M., & Kitchener, K. S. (1994). *Developing reflective judgment: Understanding and promoting intellectual growth and critical thinking in adolescents and adults*. San Francisco: Jossey Bass.
- King, P. M., & Kitchener, K. S. (2002). The reflective judgment model: Twenty years of research on epistemic cognition. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 37-62). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kinnunen, R., & Vauras, M. (1995). Comprehension monitoring and the level of comprehension in high- and low-achieving primary school children's reading. *Learning and Instruction*, 5, 143-165.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a Model of Text Comprehension and Production, *Psychological Review*, 85, 363-394.
- Kitchener, K. S. (1983). Cognition, metacognition, and epistemic cognition: A three-level model of cognitive processing. *Human Development*, 26, 106-116.
- Klayman, J. (1985). Children's decision strategies and their adaptation to task characteristics. *Organizational Behavior and Human Decision Processes*, 35, 179-201.
- Koriat, A. (1994). Memory's knowledge of its own knowledge: The accessibility account of the feeling of knowing. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing*. Cambridge, MA: The MIT Press.
- Krapp, A. (1993). Lernstrategien: Konzepte, Methoden und Befunde [Learning strategies: concepts, methods, and results]. *Unterrichtswissenschaft*, 21(4), 291-311.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory into Practice*, 41(4), 212-218.
- Krawczak, M., & Schmidtke, J. (1994). *DNA-Fingerprinting*. Heidelberg: Spektrum Akademischer Verlag GmbH.
- Krawczak, M., & Schmidtke, J. (1998). *DNA Fingerprinting. A medical perspective* (2 ed.). Oxford, UK: BIOS Scientific Publishers Ltd.
- Kreitzer, A. E., & Madaus, G. F. (1994). Empirical investigations of the hierarchical structure of the taxonomy. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's taxonomy: A forty-year retrospective* (pp. 64 - 82). Chicago: University of Chicago Press.
- Kropp, R. P., & Stoker, H. W. (1966). *The construction and validation of tests of cognitive processes as described in the taxonomy of educational objectives*. Tallahassee, FL: Florida State University.
- Kuebli, J., & Fivush, R. (1994). Children's representations and recall of event alternatives. *Journal of Experimental Child Psychology*, 58(1), 25-45.
- Kuhn, D. (1991). *Skills of argument*. Cambridge, England: Cambridge University Press.

- Kuhn, D. (2000). Metacognitive development. *Current Directions in Psychological Sciences*, 9(5), 178-181.
- Kuhn, D., Cheney, R., & Weinstock, M. (2000). The development of epistemological understanding. *Cognitive Development*, 15(3), 309-328.
- Kuhn, D., & Weinstock, M. (2002). What is epistemological thinking and why does it matter? In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 121-144). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kulbe, I. (2006). *Einfluss von Vorwissen auf das Navigationsverhalten im Hypertext: Eine qualitative Studie* [The impact of prior domain knowledge on navigational behavior: A qualitative study]. Unpublished Diploma-Thesis, Westfälische Wilhelms-Universität, Münster.
- Kunen, S., Cohen, R., & Solman, R. (1981). A levels-of-processing analysis of bloom's taxonomy. *Journal of Educational Psychology*, 73(2), 202-211.
- Landauer, T., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, 25, 259-284.
- Langer, I., Schulz von Thun, F., & Tausch, R. (1990). *Sich verständlich ausdrücken - Anleitungstexte, Unterrichtstexte, Vertragstexte, Amtstexte, Versicherungstexte, Wissenschaftstexte, u.a.* [To express oneself comprehensible – instructional texts, bureaucratic texts, etc. ] (4th ed.). München: Ernst Reinhard, GmbH & Co, Verlag.
- Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335-1342.
- Lawless, K. A., Brown, S. W., Mills, R., & Mayall, H. J. (2003). Knowledge, interest, recall and navigation: A look at hypertext processing. *Journal of Literacy Research*, 35(3), 911-934.
- Lawless, K. A., & Kulikowich, J. M. (1996). Understanding hypertext navigation through cluster analysis. *Journal of Educational Computing Research*, 14(4), 385-399.
- Leach, J., Hind, A., & Ryder, J. (2003). Designing and evaluating short teaching interventions about the epistemology of science in high school classes. *Science Education*, 87, 831-848.
- Lin, L.-M., & Zabracky, K. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology*, 23, 345-391.
- Lin, L.-M., Zabracky, K., & Moore, D. (1997). The relations among interest, self-assessed comprehension, and comprehension performance in young adults. *Reading Research and Instruction*, 36(2), 127-139.
- Lind, G., & Sandmann, A. (2003). Lernstrategien und Domänenwissen [Learning strategies and domain knowledge]. *Zeitschrift für Psychologie*, 211(4), 171-192.
- Lodewyk, K. R., & Winne, P. H. (2005). Relations among the structure of learning tasks,

- achievement, and changes in self-efficacy in secondary students. *Journal of Educational Psychology*, 97(1), 3-12.
- Luwel, K., Verschaffel, L., Onghena, P., & De Corte, E. (2003). Analyzing the adaptiveness of strategy choices using the choice / no-choice method: The case of numerosity judgments. *European Journal of Cognitive Psychology*, 15(4), 511-537.
- Luyten, L., Lowyck, J., & Tuerlinckx, F. (2001). Task perception as a mediating variable: A contribution to the validation of instructional knowledge. *British Journal of Educational Psychology*, 71, 203-223.
- Mack, W. (1996). Expertise und Intelligenz [Expertise and intelligence]. In H. Gruber & A. Ziegler (Eds.), *Expertiseforschung: Theoretische und methodische Grundlagen* (pp. 92-113). Opladen: Westdeutscher Verlag.
- Madaus, G. F., Woods, E. M., & Nuttall, R. L. (1973). A causal model analysis of Bloom's taxonomy. *American Educational Research Journal*, 10(4), 253-262.
- Magliano, J. P., Little, L. D., & Graesser, A. C. (1993). The impact of comprehension instruction on the calibration of comprehension. *Reading Research and Instruction*, 32(3), 49-63.
- Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in theory and practice* (pp. 117 – 144). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Maki, R. H., Foley, J. M., Kajer, W. K., Thompson, R. C., & Willert, M. G. (1990). Increased processing enhances calibration of comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(4), 609-616.
- Maki, R. H., & McGuire, M. J. (2002). Metacognition for text: Findings and implications for education. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition*. Cambridge: Cambridge University Press.
- Maki, R. H., & Serra, M. (1992). Role of practice tests in the accuracy of test predictions on text material. *Journal of Educational Psychology*, 84(2), 200-217.
- Maki, R. H., Shields, M., Wheeler, A. E., & Zacchilli, T. L. (2005). Individual differences in absolute and relative metacomprehension accuracy. *Journal of Educational Psychology*, 97(4), 723-731.
- Markman, E. M., & Gorin, L. (1981). Children's ability to adjust their standards for evaluating comprehension. *Journal of Educational Psychology*, 73(3), 320-325.
- Marx, R. W., & Walsh, J. (1988). Learning from academic tasks. *Elementary School Journal*, 88(3), 207-219.
- Marzano, R. J. (1992). *A different kind of classroom*. Alexandria, VA: Association of Supervision and Curriculum Development.
- Mason, L. (2003). High school students' beliefs about math, mathematical problem solving, and their achievement in math: A cross-sectional study. *Educational Psychology*, 23(1),



- 73-85.
- Mason, L., & Boscolo, P. (2004). Role of epistemological understanding and interest in interpreting a controversy and in topic-specific belief change. *Contemporary Educational Psychology, 29*, 103-128.
- Mason, L., & Scirica, F. (2006). Prediction of students' argumentation skills about controversial topics by epistemological understanding. *Learning and Instruction, 16*(5), 492-509.
- Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.
- Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into Practice, 41*(1), 226-232.
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31-48). New York: Cambridge University Press.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 43-52.
- McDonald, S., & Stevenson, R. J. (1998). Navigation in hyperspace: An evaluation of the effects of navigational tools and subject matter expertise on browsing and information retrieval in hypertext. *Interacting with Computers, 10*, 129-142.
- McEneaney, J. E. (2001). Graphic and numerical methods to assess navigation in hypertext. *International Journal of Human-Computer Studies, 55*, 761-786.
- McNamara, D. S., Kintsch, E., Songer, N., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1-43.
- McNamara, D. S., Louwerse, M. M., & Graesser, A. C. (2007). *Coh-Metrix: Automated cohesion and coherence scores to predict text readability and facilitate comprehension*. Retrieved August 27th, 2007, from <http://cohmetrix.memphis.edu/cohmetrixpr/archive/Coh-MetrixGrant.pdf>
- Merrill, M. D. (1994). *Instructional design theory*. Englewood Cliffs, NJ: Educational Technology Publications.
- Metcalfe, J. (1992). Dynamic metacognitive monitoring during problem solving. In T. O. Nelson (Ed.), *Metacognition: Core readings* (pp. 196 – 214): Allyn & Bacon.
- Metfessel, N. S., Michael, W. G., & Kirsner, D. A. (1969). Instrumentation of Bloom's and Krathwohl's taxonomies for the writing of educational objectives. *Psychology in the Schools, 6*, 227-231.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97.
- Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. New York, NY: Henry Holt and Co.
- Möller, J., & Müller-Kalthoff, T. (2000). Lernen mit Hypertext: Effekte von Navigationshilfen und Vorwissen [Learning with hypertext: Effects of navigational aids and prior

- domain knowledge]. *Zeitschrift für Pädagogische Psychologie*, 14(2/3), 116-123.
- Montalvo, F. T., & Torres, M. C. G. (2004). Self-regulated learning: Current and future directions. *Electronic Journal of Research in Educational Psychology*, 2(1), 1-34.
- Moore, R. W., & Foy, R. L. (1997). The scientific attitude inventory: A revision (SAI II). *Journal of Research in Science Teaching*, 34(4), 327-336.
- Moore, W. S. (2002). Understanding learning in a postmodern world: Reconsidering the Perry scheme of ethical and intellectual development. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 17-36). Mahwah, NJ: Lawrence Erlbaum Associates.
- Moos, D., & Azevedo, R. (2006). Examining the fluctuation of strategy use during learning with hypermedia. In S. A. Barab, K. E. Hay & D. T. Hickey (Eds.), *Proceedings of the 7th International Conference of the Learning Sciences* (pp. 481-487). Mahwah, NJ: Erlbaum.
- Morris, C. C. (1990). Retrieval processes underlying confidence in comprehension judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(2), 223-232.
- Murray, T. (2003). MetaLinks: Authoring and affordances for conceptual and narrative flow in adaptive hyperbooks. *Journal of Artificial Intelligence in Education*, 13, 199-233.
- National Research Council. (1996). *The evaluation of forensic DNA evidence*. Washington, D.C.: National Academy Press.
- Neber, H., & Schommer-Aikins, M. (2002). Self-regulated science learning with highly gifted students: The role of cognitive, motivational, epistemological, and environmental variables. *High Ability Studies*, 13(1), 59-74.
- Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin*, 95(1), 109-133.
- Nelson, T. O. (1996). Gamma is a measure of the accuracy of predicting performance on one item relative to another item, not of the absolute performance on an individual item. Comments on Schraw (1995). *Applied Cognitive Psychology*, 10, 257-260.
- Nelson, T. O. (1999). Cognition versus metacognition. In R. J. Sternberg (Ed.), *The nature of cognition* (pp. 625-641): MIT Press.
- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect". *Psychological Science*, 2(4), 267-270.
- Nelson, T. O., & Narens, L. (1994). Why investigate metacognition. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition, knowing about knowing* (pp. 1-25). Cambridge: MIT Press.
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. New York: Prentice-Hall.
- Niessen, T., Vermunt, J. D., Abma, T., Widdershoven, G., & van der Vleuten, C. (2004).

- On the nature and form of epistemologies: Revealing hidden assumptions through an analysis of instrument design. *European Journal of School Psychology*, 2(1-2), 39-64.
- Nietfeld, J., & Schraw, G. (2002). The effect of knowledge and strategy training on monitoring accuracy. *The Journal of Educational Research*, 95(3), 131-142.
- Nietfeld, J. L., Enders, C. K., & Schraw, G. (2006). A monte carlo comparison of measures of relative and absolute monitoring accuracy. *Educational and Psychological Measurement*, 66(2), 258-271.
- Ormell, C. P. (1975). Bloom's taxonomy and the objectives of education. *Educational Research*, 17, 3-18.
- Otero, J. (1998). Influence of knowledge activation and context on comprehension monitoring of science texts. In D. J. Hacker, J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in theory and practice* (pp. 145 – 164). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Otero, J. (2002). Noticing and fixing difficulties while understanding science texts. In J. Otero, J. A. León & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 281 – 307). Mahwah, NJ: Lawrence Erlbaum Associates.
- Otero, J., & Campanario, J. M. (1992). The relationship between academic achievement and metacognitive comprehension monitoring ability of Spanish secondary school students. *Educational and Psychological Measurement*, 52, 419-430.
- Paulsen, M. B., & Wells, C. T. (1998). Domain differences in the epistemological beliefs of college students. *Research in Higher Education*, 39(4), 365-384.
- Payne, J. W. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Behavior & Human Performance*, 16(2), 366-387.
- Perry, N. E., Phillips, L., & Dowler, J. (2004). Examining features of tasks and their potential to promote self-regulated learning. *Teachers College Record*, 106(9), 1854-1878.
- Perry, W. G. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart, and Winston.
- Pfeiffer, H., & Brinkmann, B. (2000). Sequenzanalyse mitochondrialer DNA in der Rechtsmedizin [Sequence analysis of mitochondrial DNA in forensic medicine]. *Rechtsmedizin*, 10, 51-55.
- Piaget, J. (1985). *The equilibration of cognitive structures*. Chicago: University of Chicago Press.
- Pieschl, S., Stahl, E., & Bromme, R. (2006). Impact of epistemological beliefs on hypertext learning: A pilot study. In R. Sun & N. Miyake (Eds.) *Proceedings of The 28<sup>th</sup> Annual Conference of the Cognitive Science Society (CogSci)*, Vancouver, Canada (pp. 1956 – 1961).
- Pieschl, S., Stahl, E., & Bromme, R. (2008). Epistemological beliefs and self-regulated learning with hypertext. *Metacognition and Learning*, 1, 17-37.

- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451 – 502). San Diego, CA.
- Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice*, 41(4), 219-225.
- Pintrich, P. R. (2002). Future challenges and directions for theory and research on personal epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 389-414). Mahwah, NJ: Lawrence Erlbaum Associates.
- Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385-407.
- Pintrich, P. R., Wolters, C. A., & Baxter, G. P. (2000). Assessing metacognition and self-regulated learning. In G. Schraw & J. C. Impara (Eds.), *Issues in the measurement of metacognition* (pp. 43 – 97). Lincoln, Nebraska: Buros Institute of Mental Measurements.
- Plumb, C., Butterfield, E. C., Hacker, D. J., & Dunlosky, J. (1994). Error correction in text. *Reading and Writing*, 6(4), 347-360.
- Potelle, H., & Rouet, J.-F. (2003). Effects of content representation and readers' prior knowledge on the comprehension of hypertext. *International Journal of Human-Computer Studies*, 58, 327-345.
- Pressley, M., & Ghatala, E. S. (1988). Delusions about performance on multiple-choice comprehension tests. *Reading Research Quarterly*, 24(4), 454-464.
- Pressley, M., & Ghatala, E. S. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist*, 25(1), 19-33.
- Pressley, M., Yokoi, L., van Meter, P., Van Etten, S., & Freebern, G. (1997). Some of the reasons why preparing for exams is so hard: What can be done to make it easier? *Educational Psychology Review*, 9(1), 1-38.
- Priest, A. G., & Lindsay, R. O. (1992). New light on novice-expert differences in physics problem solving. *British Journal of Psychology*, 83, 389-405.
- Puers, C. (2003). *DNA-Analyse in der Kriminaltechnik* [DNA analysis in criminal techniques]. Retrieved 13. September, 2003, from <http://www.uni-tuebingen.de/ZMBP/life/Kriminaltechnik.pdf>
- Puustinen, M., & Pulkkinen, L. (2001). Models of self-regulated learning: A review. *Scandinavian Journal of Educational Research*, 45(3), 269-286.
- Qian, G., & Alvermann, D. (1995). Role of epistemological beliefs and learned helplessness in secondary school students' learning science concepts from text. *Journal of Educational Psychology*, 87(2), 282-292.
- Qian, G., & Alvermann, D. (2000). Relationship between epistemological beliefs and con-

- ceptual change learning. *Reading & Writing Quarterly*, 16, 59-74.
- Quellmalz, E. (1987). Developing reasoning skills. In J. B. Faron & R. J. Sternberg (Eds.), *Teaching thinking skills* (pp. 86-105). New York: W.H. Freeman.
- Quereshi, M. Y. (1989). Logical versus empirical readability and human interest scores for introductory texts published between 1977 and 1988. *Teaching of Psychology*, 16(1), 4-7.
- Rabinowitz, M., Freeman, K., & Cohen, S. (1992). Use and maintenance of strategies: The influence of accessibility to knowledge. *Journal of Educational Psychology*, 84(2), 211-218.
- Rawson, K. A., Dunlosky, J., & Thiede, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory & Cognition*, 28(6), 1004-1010.
- Reid, N., & Yang, M.-J. (2002). The solving of problems in chemistry: The more open-ended problems. *Research in Science & Technological Education*, 20(1), 83-98.
- Reigeluth, C. M., & Moore, J. (1999). Cognitive education and the cognitive domain. In C. M. Reigeluth (Ed.), *Instructional-design theories and models, Vol. II: A new paradigm of instructional theory*. Mahwah, NJ: Erlbaum.
- Rellinger, E. R., Borkowski, J. G., Turner, L. A., & Hale, C. A. (1995). Perceived task difficulty and intelligence: determinants of strategy use and recall. *Intelligence*, 20, 125-143.
- Reynolds, R. E. (1992). Selective attention and prose learning: Theoretical and empirical research. *Educational Psychology Review*, 4(4), 345-391.
- Richter, T., Naumann, J., & Noller, S. (2003). LOGPAT: A semi-automatic way to analyze hypertext navigation behavior. *Swiss Journal of Psychology*, 62(2), 113-120.
- Rikers, R. M. J. P., & Paas, F. (2005). Recent advances in expertise research. *Applied Cognitive Psychology*, 19, 145-149.
- Rikers, R. M. J. P., Schmidt, H. G., & Boshuizen, H. P. A. (2000). Knowledge encapsulation and the intermediate effect. *Contemporary Educational Psychology*, 25, 150-166.
- Rikers, R. M. J. P., Schmidt, H. G., & Moulart, V. (2005). Biomedical knowledge: Encapsulated or two worlds apart? *Applied Cognitive Psychology*, 19, 223-231.
- Rohwer, W. D., & Sloane, K. (1994). Psychological perspectives. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's taxonomy: A forty-year retrospective* (pp. 41 - 64). Chicago: University of Chicago Press.
- Rolf, B., & Wiegand, P. (2004). Analyse biologischer Spuren. Teil III: Mitochondriale DNA und Y-chromosomale STR [Analysis of biological traces. Part III: Mitochondrial DNA and Y-chromosomal STR]. *Rechtsmedizin*, 14, 473-484.
- Romizowski, A. J. (1981). *Designing instructional systems: Decision making in course planning and curriculum design*. London: Kogan Page.

- Rothe, H.-J., & Schindler, M. (1996). Expertise und Wissen [Expertise and knowledge]. In H. Gruber & A. Ziegler (Eds.), *Expertiseforschung: Theoretische und methodische Grundlagen* (pp. 35-57). Opladen: Westdeutscher Verlag.
- Rouet, J.-F. (2003). What was I looking for? The influence of task specificity and prior knowledge on students' search strategies in hypertext. *Interacting with Computers*, 15, 409-428.
- Rouet, J.-F., Favart, M., Britt, M. A., & Perfetti, C. A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction*, 15(1), 85-106.
- Rouet, J.-F., & Tricot, A. (1996). Task and activity models in hypertext usage. In H. van Oostendorp & S. de Mul (Eds.), *Cognitive aspects of electronic text processing* (pp. 239-264). Norwood, NJ: Ablex.
- Rouet, J.-F., Vidal-Abarca, E., Erbou, A. B., & Millogo, V. (2001). Effects of information search tasks on the comprehension of instructional text. *Discourse Processes*, 31(2), 163-186.
- Rozendaal, J. S., de Brabander, C. J., & Minnaert, A. E. (2001). *Boundaries and dimensionality of epistemological beliefs*. Paper presented at the bi-annual conference of the European Association of Research on Learning and Instruction (EARLI), Fribourg, Switzerland.
- Ryan, M. P. (1984). Monitoring text comprehension: Individual differences in epistemological standards. *Journal of Educational Psychology*, 76(2), 248-258.
- Salas, A., Lareu, M. V., & Carracedo, A. (2001). Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: A case report. *International Journal of Legal Medicine*, 114, 186-190.
- Salmerón, L., Kintsch, W., & Canas, J. (2006). Reading strategies and prior knowledge in learning from hypertext. *Memory & Cognition*, 34(5), 1157-1171.
- Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education*, 89(4), 634-656.
- Sandoval, W. A., & Morrison, K. (2003). High school students' ideas about theories and theory change after a biological inquiry unit. *Journal of Research in Science Teaching*, 40(4), 369-392.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993a). On acquiring expertise in medicine. *Educational Psychology Review*, 5(3), 205-221.
- Schmidt, H. G., & Boshuizen, H. P. A. (1993b). On the origin of intermediate effects in clinical case recall. *Memory & Cognition*, 21(3), 338-351.
- Schmitter, H. (1998). DNA-Identifizierungsmuster in der Strafverfolgung [DNA identification patterns in prosecution]. *Spektrum der Wissenschaft*, 10, 56-59.
- Schneider, W., Schlagmüller, M., & Visé, M. (1998). The impact of metamemory and do-

- main-specific knowledge on memory performance. *European Journal of Psychology of Education*, 13(1), 91-103.
- Schnotz, W. (2005). An integrated model of text picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia* (pp. 49-69). New York: Cambridge University Press.
- Schoenfeld, A. H. (1989). Explorations of students' mathematical beliefs and behavior. *Journal for Research in Mathematics Education*, 20(4), 338-355.
- Schommer, M. (1990). Effects of beliefs about the nature of knowledge on comprehension. *Journal of Educational Psychology*, 82(3), 498-504.
- Schommer, M. (1993). Epistemological development and academic performance among secondary students. *Journal of Educational Psychology*, 85(3), 406-411.
- Schommer, M. (1994). An emerging conceptualization of epistemological beliefs and their role in learning. In R. Garner & P. A. Alexander (Eds.), *Beliefs about text and instruction with text* (pp. 25-40). Hillsdale: Lawrence Erlbaum Associates.
- Schommer, M., Calvert, C., Gariglietti, G., & Bajaj, A. (1997). The development of epistemological beliefs among secondary students: A longitudinal study. *Journal of Educational Psychology*, 89(1), 37-40.
- Schommer, M., Crouse, A., & Rhodes, N. (1992). Epistemological beliefs and mathematical text comprehension: Believing it is simple doesn't make it so. *Journal of Educational Psychology*, 84(4), 435-443.
- Schommer, M., & Surber, J. R. (1986). Comprehension-monitoring failure in skilled adult readers. *Journal of Educational Psychology*, 78(5), 353-357.
- Schommer, M., & Walker, K. (1995). Are epistemological beliefs similar across domains? *Journal of Educational Psychology*, 87(3), 424-232.
- Schommer-Aikins, M. (2002). An evolving theoretical framework for an epistemological belief system. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 103-118). Mahwah, NJ: Lawrence Erlbaum Associates.
- Schommer-Aikins, M. (2004). Explaining the epistemological belief system: Introducing the embedded systemic model and coordinated research approach. *Educational Psychologist*, 39(1), 19-29.
- Schommer-Aikins, M., & Hutter, R. (2002). Epistemological beliefs and thinking about everyday controversial issues. *The Journal of Psychology*, 136(1), 5-20.
- Schoske, R., Vallone, P. M., Kline, M. C., Redman, J. W., & Butler, J. M. (2004). High-throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays. *Forensic Science International*, 139, 107-121.
- Schraw, G. (1995). Measures of feeling-of-knowing accuracy: A new look at an old problem. *Applied Cognitive Psychology*, 9, 321-332.
- Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*, 26, 113-

- 125.
- Schraw, G. (2001). Current themes and future directions in epistemological research: A commentary. *Educational Psychology Review*, 13(4).
- Schraw, G., Bendixen, L. D., & Dunkle, M. E. (2002). Development and validation of the epistemic belief inventory. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 261-276). Mahwah, NJ: Lawrence Erlbaum Associates.
- Schraw, G., Dunkle, M. E., & Bendixen, L. D. (1995). Cognitive processes in well-defined and ill-defined problem solving. *Applied Cognitive Psychology*, 9(6), 523-538.
- Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review*, 7(4), 351-371.
- Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. *Contemporary Educational Psychology*, 18(4), 445-463.
- Schraw, G., & Roedel, T. D. (1994). Test difficulty and judgment bias. *Memory & Cognition*, 22(1), 63-69.
- Schunn, C. D., Lovett, M. C., & Reder, L. (2001). Awareness and working memory in strategy adaptivity. *Memory & Cognition*, 29(2), 254-266.
- Seddon, G. M. (1978). The properties of Bloom's taxonomy of educational objectives for the cognitive domain. *Review of Educational Research*, 48(2), 303-323.
- Simon, D. P., & Simon, H. A. (1978). Individual differences in solving physics problems. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp. 325-348). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sinatra, G. M., Southerland, S. A., McConaughy, F., & Demastes, J. W. (2003). Intentions and beliefs in students' understanding and acceptance of biological evolution. *Journal of Research in Science Teaching*, 40(5), 510-528.
- Smith, R. B. (1968). An empirical examination of the assumptions underlying the 'taxonomy of educational objectives for the cognitive domain'. *Journal of Educational Measurement*, 5(1), 125-127.
- Smith, M. U. (1990). *Problem categorization and expertise: An experimental study of predicted differences among genetics faculty, counselors, and students*. Paper presented at the Annual Meeting of the American Educational Research Association (AERA), Boston, MA.
- Snow, C. (2002). *Reading for understanding: Toward an R & D program in reading comprehension*. Santa Monica: RAND.
- Son, L. K., & Schwartz, B. L. (2002). The relation between metacognitive monitoring and control. In T. J. Perfect & B. L. Schwartz (Eds.), *Applied metacognition* (pp. 15 - 38). Cambridge: Cambridge University Press.
- Songer, N. B., & Linn, M. C. (1991). How do students' views of science influence knowledge integration? *Journal of Research in Science Teaching*, 28(9), 761-784.
- Spencer, C. A. (2004). *Genetic testimony. A guide to forensic DNA profiling*. Upper Saddle River,



- NJ: Pearson Education Inc.
- Spiro, R. J., Feltovich, P. J., & Jacobson, M. J. (1992). Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. In T. M. Duffy & D. H. Jonassen (Eds.), *Constructivism and the technology of instruction: A conversation* (pp. 57-75). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Spörer, N., & Brunstein, J. C. (2005). Diagnostik von selbstgesteuertem Lernen. Ein Vergleich zwischen Fragebogen- und Interviewverfahren [Diagnosis of self-regulated learning. A comparison between questionnaires and interviews]. In B. Moschner & C. Artelt (Eds.), *Lernstrategien und Metakognition: Implikationen für Forschung und Praxis* (pp. 43 – 63). Münster: Waxmann.
- Spurlin, J. E., Dansereau, D. F., & O'Donnell, A. (1988). Text processing: Effects of summarization frequency on text recall. *Journal of Experimental Education*, 56(4), 199-202.
- Stadtler, M. (2006). *Auf der Suche nach medizinischen Fachinformationen. Metakognitionen bei der Internetrecherche von Laien* [Searching for medical information: Metacognitions during internet search of novices]. Münster: Waxman.
- Stadtler, M., & Bromme, R. (2005). *Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models*. Paper presented at the 27th Annual Conference of the Cognitive Science Society (CogSci), Stresa, Italy.
- Stahl, E., & Bromme, R. (in press). CAEB. An instrument to measure connotative aspects of epistemological beliefs. *Learning and Instruction*.
- Stahl, E., Bromme, R., Stadtler, M., & Jaron, R. (2006). Learning by hypertext writing: Effects of considering a single audience versus multiple audiences on knowledge acquisition. In D. Galbraith, L. van Waes & M. Torrance (Eds.), *Recent developments in writing process research* (Vol. 2: Methods and applications, pp. 307-321). Dordrecht: Kluwer.
- Stahl, E., Pieschl, S., & Bromme, R. (2006). Task complexity, epistemological beliefs and metacognitive calibration: An exploratory study. *Journal of Educational Computing Research*, 35(4), 319-338.
- Stahl, R. J. (1979). *The domain of cognition: An alternative to Bloom's taxonomy within the framework of an information processing model of learning*. Paper presented at the Annual Meeting of the Rocky Mountain Educational Research Association, Tucson, Arizona.
- Sternberg, R. J. (1997). Cognitive conceptions of expertise. In P. J. Feltovich, K. M. Ford & R. R. Hoffman (Eds.), *Expertise in context: human and machine* (pp. 149-161). Menlo Park: American Association for Artificial Intelligence.
- Sternberg, R. J. (1998). Metacognition, abilities, and developing expertise: What makes an expert student? *Instructional Science*, 26, 127-140.
- Stevens, J. P. (1996). *Applied multivariate statistics for the social sciences* (3 ed.). New Jersey: Lawrence Erlbaum Associates.

- Stodolsky, S., Salk, S., & Glaessner, B. (1991). Student views about learning math and social studies. *American Educational Research Journal*, 28(1), 89-116.
- Stone, N. J. (2000). Exploring the relationship between calibration and self-regulated learning. *Educational Psychology Review*, 12(4), 437-475.
- Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). New York: Cambridge University Press.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251-296.
- Szibor, R., Michael, M., Plate, I., & Krause, D. (2000). Efficiency of forensic mtDNA analysis. Case examples demonstrating the identification of traces. *Forensic Science International*, 113, 71-78.
- Tabachnik, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5 ed.). Boston: Pearson.
- Thiede, K. W., & Anderson, M. C. M. (2003). Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology*, 28, 129-160.
- Thiede, K. W., & Dunlosky, J. (1999). Towards a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology-Learning Memory and Cognition*, 25(4), 1024-1037.
- Thomas, E. L., & Robinson, H. A. (1982). *Improving reading in every class: A sourcebook for teachers*. Boston: Allyn & Bacon.
- Tolhurst, D. (2004). *The influence of web-supported independent activities and small group work on students' epistemological beliefs*. Paper presented at the Sixth Australasian Computing Education Conference (ACE2004), Dunedin, New Zealand.
- Trautwein, U., & Lüdke, O. (in press). Predicting global and topic-specific certainty beliefs: Domain-specificity and the role of academic environment. *British Journal of Educational Psychology*.
- Trautwein, U., Lüdke, O., & Beyer, B. (2004). Rauchen ist tödlich, Computerspiele machen aggressiv? Allgemeine und theorienspezifische epistemologische Überzeugungen bei Studierenden unterschiedlicher Fachrichtungen [Smoking kills, computer games make aggressive? General and theory-specific epistemological beliefs of students with different majors]. *Zeitschrift für Pädagogische Psychologie*, 18(3-4), 187-199.
- Tsai, C.-C. (1998). An analysis of scientific epistemological beliefs and learning orientations of Taiwanese eighth graders. *Science Education*, 82(4), 473-489.
- Tully, G., Bär, W., Brinkmann, B., Carracedo, A., Gill, P., Morling, N., et al. (2001). Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. *Forensic Science International*, 124, 83-91.
- Tuovinen, J., & Sweller, J. (1999). A comparison of cognitive load associated with discovery

- learning and worked examples. *Journal of Educational Psychology*, *91*, 334-341.
- Urhahne, D., & Hopf, M. (2004). Epistemologische Überzeugungen in den Naturwissenschaften und ihre Zusammenhänge mit Motivation, Selbstkonzept und Lernstrategien [Epistemological beliefs in sciences and their relations to motivation, self-concept and learning strategies]. *Zeitschrift für Didaktik der Naturwissenschaften*, *10*, 70-86.
- Valanides, N., & Angeli, C. (2005). Effects of instruction on changes in epistemological beliefs. *Contemporary Educational Psychology*, *30*, 314-330.
- Van de Watering, G., & van der Rijt, J. (2006). Teachers' and students' perceptions of assessments: A review and a study into the ability and accuracy of estimating the difficulty level of assessment items. *Educational Research Review*, *1*, 133-147.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data. *Applied Cognitive Psychology*, *19*, 205-221.
- Van Meter, P., Yokoi, L., & Pressley, M. (1994). College students' theory of note-taking derived from their perceptions of note-taking. *Journal of Educational Psychology*, *86*(3), 323-338.
- Veenman, M., & Beishuizen, J. (2004). Intellectual and metacognitive skills of novices while studying texts under conditions of text difficulty and time constraints. *Learning and Instruction*, *14*(6), 621-640.
- Veenman, M., & Elshout, J. J. (1999). Changes in the relationship between cognitive and metacognitive skills during the acquisition of expertise. *European Journal of Psychology of Education*, *XIV*(4), 509-523.
- Veenman, M. V. J. (2003). *The assessment of metacognitive skills: What can be learned from multi-method designs?* Paper presented at the bi-annual conference of the European Association of Research on Learning and Instruction (EARLI), Padua, Italy.
- Wagoner, S. A. (1983). Comprehension monitoring: What it is and what we know about it. *Reading Research Quarterly*, *18*, 328-346.
- Walczyk, J. J., & Hall, V. C. (1989). Is the failure to monitor comprehension an instance of cognitive impulsivity? *Journal of Educational Psychology*, *81*(3), 294-298.
- Watt, H. M. G. (2004). Development of adolescents' self-perception, values, and task perception according to gender and domain in 7th- through 11th-grade Australian students. *Child Development*, *75*(5), 1556-1574.
- Weaver, C. A., & Bryant, D. S. (1995). Monitoring of comprehension: The role of text difficulty in metamemory for narrative and expository text. *Memory & Cognition*, *23*(1),

- 12-22.
- Weber, T. P. (2002). *Schnellkurs Genforschung* [A short course in genetic research]. Köln: DuMont.
- Weingardt, K. R., Leonesio, R. J., & Loftus, E. F. (1994). Viewing eyewitness research from a metacognitive perspective. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition, knowing about knowing* (pp. 157 – 184). Cambridge: MIT Press.
- Whitmire, E. (2003). Epistemological beliefs and the information-seeking behavior of undergraduates. *Library & Information Science Research*, 25, 127-142.
- Wiegand, P., & Rolf, B. (2003). Analyse biologischer Spuren. Teil II: DNA-Typisierung [Analysis of biological traces. Part II: DNA-typing]. *Rechtsmedizin*, 13, 375-383.
- Wild, K.-P. (1998). Lernstrategien und Lernstile [Learning strategies and learning styles]. In D. E. Rost (Ed.), *Handwörterbuch der Pädagogischen Psychologie* (pp. 309-312). Weinheim: Beltz.
- Wild, K.-P., & Schiefele, U. (1994). Lernstrategien im Studium. Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens [Learning strategies during studying. Results with regard to factor structure and reliability of a new questionnaire]. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 15, 185-200.
- Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology*, 132(4), 408-428.
- Williams, R. G. (1977). A behavioral typology of educational objectives for the cognitive domain. *Educational Technology*, 17(6), 36-46.
- Windschitl, M., & Andre, T. (1998). Using computer simulations to enhance conceptual change: The roles of constructivist instruction and student epistemological beliefs. *Journal of Research in Science Teaching*, 35(2), 145-160.
- Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational Psychologist*, 30(4), 173-187.
- Winne, P. H. (1996). A metacognitive view of individual differences in self-regulated learning. *Learning and Individual Differences*, 8(4), 327-353.
- Winne, P. H. (1997). Experimenting to bootstrap self-regulated learning. *Journal of Educational Psychology*, 89(3), 397-410.
- Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement. Theoretical perspectives* (pp. 153-189). Mahwah, NJ: Lawrence Erlbaum Associates.
- Winne, P. H. (2005). Key issues in modeling and applying research on self-regulated learning. *Applied Psychology: An International Review*, 54(2), 232-238.
- Winne, P. H. (2006). How software technologies can improve research on learning and bolster school reform. *Educational Psychologist*, 41(1), 5-17.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker,

- J. Dunlosky & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277-304). Mahwah, NJ: Lawrence Erlbaum Associates.
- Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology, 27*(4), 551-572.
- Winne, P. H., & Jamieson-Noel, D. (2003). Self-regulating studying by objectives for learning: Students' reports compared to a model. *Contemporary Educational Psychology, 28*, 259-276.
- Winne, P. H., Jamieson-Noel, D., & Muis, K. R. (2002). Methodological issues and advances in researching tactics, strategies, and self-regulated learning. In P. R. Pintrich & M. L. Maehr (Eds.), *New directions in measures and methods* (pp. 121 – 155). Oxford, UK: Elsevier Sciences.
- Winne, P. H., & Marx, R. W. (1982). Students' and teachers' views of thinking processes for classroom learning. *Elementary School Journal, 82*(5), 493-518.
- Winne, P. H., & Marx, R. W. (1989). A cognitive processing analysis of motivation within classroom tasks. In C. Ames & R. Ames (Eds.), *Research on motivation in education* (pp. 223-257). Orlando, FL: Academic Press.
- Winne, P. H., & Muis, K. R. (2004). *Statistical estimates of learner's judgments about knowledge in calibration of achievement*. Unpublished manuscript, Simon Fraser University, Burnaby, BC, Canada.
- Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. R. Pintrich & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 531-566). San Diego: Academic Press.
- Winograd, P., & Johnston, P. (1982). Comprehension monitoring and the error detection paradigm. *Journal of Reading Behavior, XIV*(1), 61-76.
- Wittwer, J., Bromme, R., & Jucks, R. (2004). Kann man dem Internet trauen, wenn es um die Gesundheit geht? Die Glaubwürdigkeitsbeurteilung medizinischer Fachinformation im Internet durch Laien [Can you trust the internet in health issues? Evaluation of trustworthiness of medical information on the internet by novices]. *Zeitschrift für Medienpsychologie, 2*, 48-56.
- Wolters, C. A., Yu, S. L., & Pintrich, P. R. (1996). The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learning and Individual Differences, 8*(3), 211-238.
- Wood, P., & Kardash, C. M. (2002). Critical elements in the design and analysis of studies of epistemology. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology. The psychology of beliefs about knowledge and knowing* (pp. 231-260). Mahwah, NJ: Lawrence Erlbaum Associates.
- Yeung, A., Jin, P., & Sweller, J. (1998). Cognitive load and learner expertise: Split-attention

- and redundancy effects in reading with explanatory notes. *Contemporary Educational Psychology*, 23, 1-21.
- Zabrucky, K., & Moore, D. (1989). Children's ability to use three standards to evaluate their comprehension of text. *Reading Research Quarterly*, 24, 336-352.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35(2), 151-175.
- Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 81(3), 329-339.
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3-17.
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: an overview. *Theory into Practice*, 41(2), 64-70.
- Zimmerman, B. J. (2006). Development and adaptation of expertise. In K. A. Ericsson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *Cambridge handbook of expertise and expert performance* (pp. 705-722). New York: Cambridge University Press.

## Appendix

### Appendix<sup>9</sup> A: Study I

#### Online Session

- A1 cover page (G / E)
- A2 CAEB (Connotative Aspects of Epistemological Beliefs) (G / E)
- A3 EBI (Epistemological Beliefs Inventory) (G / E)
- A4 demographic questions (G / E)

#### Face to Face Session

- A5 *prior domain knowledge test (molecular biology)* (G / E) ..... 255
- A6 neutral introduction to genetic fingerprinting (G)
- A7 epistemological introduction to genetic fingerprinting (G)
- A8 CAEB (Connotative Aspects of Epistemological Beliefs) (G / E)
- A9 *COPEs-questionnaire* (G / E) ..... 257
- A10 *tasks representing the six Bloom-Categories* (G / E) ..... 265

### Appendix B: Study II

#### Online Session

- B1 cover page (G / E)
- B2 CAEB (Connotative Aspects of Epistemological Beliefs) (G / E)
- B3 GCBS (General Certainty Beliefs Scale) (G / E)
- B4 demographic questions (G / E)

#### Face to Face Session

- B5 prior domain knowledge test (molecular biology) (G / E) (see A5)
- B6 revised neutral introduction to genetic fingerprinting (G)

---

<sup>9</sup> This chapter only includes the English versions of the most important materials specifically developed for these studies (*in italics*). For a thorough overview of all listed appendices, see the corresponding DVD-ROM including all non-originally developed materials (e.g., epistemological beliefs questionnaires) as well as the original German versions. Most materials are available in German and in English with few exceptions (e.g., the whole hyperbook on “genetic fingerprinting” was not translated into English). The available languages are indicated by capitalized letters in brackets after each appendix (G for German and E for English). In the DVD-ROM version clicking on these letters opens the corresponding document. The DVD-ROM is not appended in this online version. Please contact the author if you need an appendix (pieschl@uni-muenster.de).

B7	revised epistemological introduction to genetic fingerprinting (G)	
B8	CAEB (Connotative Aspects of Epistemological Beliefs) (G / E) (see A8)	
B9	GCBS (General Certainty Beliefs Scale) (G / E)	
B10	hyperbook on genetic fingerprinting (readme E, hyperbook G)	
B11	instruction about the structure and navigation of the hyperbook (G)	
B12	diagram of a hierarchical structure (diagram)	
B13	diagram of the overall structure of the hypertext (diagram)	
B14	diagram of the hypertext structure of STR analysis (G)	
B15	list of navigational commands (G)	
B16	instruction about task solution and prompting (G)	
B17	<i>tasks representing different Bloom-Categories</i> (G / E) .....	267
B18	interviewer guide for the retrospective stimulated recall interview (G)	
B19	coding scheme for the concurrent thoughts elicited by prompts (G)	
B20	coding scheme for the retrospective stimulated recall interviews (G)	

### Appendix C: Study III

#### Face to Face Session

C1	cover page and demographic questions (G / E)	
C2	WKI (Wood and Kardash Instrument) (G / E)	
C3	CAEB (Connotative Aspects of Epistemological Beliefs) (G / E)	
C4	prior domain knowledge test (molecular biology) (G / E) (see A5)	
C5	hyperbook on mtDNA analysis (readme E, hyperbook G)	
C6	instruction about the structure and navigation of the hyperbook (G)	
C7	diagram of the hypertext structure of mtDNA analysis (G)	
C8	list of “related information” nodes (G)	
C9	list of navigational commands (G) (see B15)	
C10	instruction for learning and comprehensibility questionnaire (G / E)	
C11	<i>learning outcome test</i> (G / E) .....	269



## Prior Domain Knowledge in Genetics<sup>10</sup>

The following page contains some multiple-choice question from the domain of genetics. Partly, these questions refer directly to the topic of “genetic fingerprinting”. These questions measure your prior domain knowledge. Please, answer all questions. If you don’t know the answer to a question, please check the category “I don’t know”.

1	What does the abbreviation DNA stand for?	<input type="radio"/> Dinatriumacetat <input type="radio"/> Deoxyribonucleic acid <input type="radio"/> Dodecyl nitric acid <input type="radio"/> Deoxy natrium acid <input type="radio"/> Dodecyl nitric acetat <input type="radio"/> I don't know
2	Where in a healthy animal cell do you find DNA?	<input type="radio"/> in the cell nucleus <input type="radio"/> in the cell nucleus and the mitochondria <input type="radio"/> in the cellular fluid <input type="radio"/> in the ribosoms <input type="radio"/> in the cellular fluid and in the mitochondria <input type="radio"/> I don't know
3	Which of these substances is contained in pure DNA?	<input type="radio"/> metal ions <input type="radio"/> sugar <input type="radio"/> phospholipides <input type="radio"/> amino acids <input type="radio"/> sulfur <input type="radio"/> I don't know
4	How is DNA structured?	<input type="radio"/> linear chain <input type="radio"/> chain with intersections <input type="radio"/> combination of $\alpha$ -helices and $\beta$ -structures <input type="radio"/> a single multi-choin strand <input type="radio"/> double helix <input type="radio"/> I don't know

<sup>10</sup> In the original questionnaire, all questions (1 – 8) fit on one page. Although the original format was retained (e.g., same fonts), the page settings differed between the original questionnaire and this thesis. Thus, the factual questions are distributed across two pages.

5	What does the "dogma of molecular biology" say?	<ul style="list-style-type: none"><li><input type="radio"/> ribosom → RNA → protein</li><li><input type="radio"/> DNA → mitochondria → protein</li><li><input type="radio"/> DNA + RNA = protein</li><li><input type="radio"/> DNA → RNA → protein</li><li><input type="radio"/> DNA + amino acids = protein</li><li><input type="radio"/> I don't know</li></ul>
6	How do you separate DNA fragments?	<ul style="list-style-type: none"><li><input type="radio"/> gel electrophoresis</li><li><input type="radio"/> synchronous fertilization</li><li><input type="radio"/> with iso-electric focusing</li><li><input type="radio"/> asynchronous fertilization</li><li><input type="radio"/> heating of the fragments</li><li><input type="radio"/> I don't know</li></ul>
7	What does the abbreviation PCR stand for?	<ul style="list-style-type: none"><li><input type="radio"/> Protein Coupling Reaction</li><li><input type="radio"/> Phosphate Chain Reaction</li><li><input type="radio"/> Polymerase Chain Reaction</li><li><input type="radio"/> Polysaccharide Chain Reaction</li><li><input type="radio"/> Phosphate Coupling Reaction</li><li><input type="radio"/> I don't know</li></ul>
8	What do you do with a PCR?	<ul style="list-style-type: none"><li><input type="radio"/> sort DNA fragments according to size</li><li><input type="radio"/> isolate DNA from human tissue</li><li><input type="radio"/> determine loci for DNA analysis</li><li><input type="radio"/> exponential amplification of DNA fragments</li><li><input type="radio"/> coloring of DNA bands</li><li><input type="radio"/> I don't know</li></ul>



WWU Münster

**Educational Psychology**  
**Psychological Institute III**  
**Research group Prof. Dr. R. Bromme**

## COPES-Questionnaire

Dear Student,

In the main part of this study you will evaluate six tasks from the topic of “genetic fingerprinting” that we developed recently. You don’t have to solve these tasks! For each task you will answer multiple questions. The task you have to evaluate will always be presented on the left side of the questionnaire. The questions on the right side always pertain to the corresponding task on the left side. Please don’t be discouraged by the number of questions. Imagine for each question that you had to solve the task with the help of our new learning environment. This learning environment is conceptualized like a hypermedia book explaining the biological background and the methods involved in “genetic fingerprinting” in detail. To solve these tasks, you would be able to search for information in this “book”.

The questions in the first part pertain to *conditions* that should be given according to your opinion to solve the task correctly. In the second part, you have to judge what kind of *learning strategies* you would enact during task solution, the third part pertains to *goals or standards* that you would set yourself for task solution and the fourth part requires an overall *evaluation* of these tasks.

We are interested in your personal opinion. There are no right or wrong answers.

Thank you for your cooperation!!!

In order to facilitate matching of the two questionnaires, please fill in a personal code in the adequate fields in both questionnaires. Your personal code consists of:

**1st character:** First letter of your mother’s first name.

**2nd character:** First letter of your father’s first name.

**3rd character:** First letter of your own first name.

**4th character:** First letter of your place of birth.

**5th and 6th character:** DOB (day of birth).

(e.g. birthday: 12th of September → The 5th character is a „1“ and the 6th character a „2“;

e.g. birthday: 02nd of September → The 5th character is a „0“ and the 6th character a „2“)

Personal Code:

1.	2.	3.	4.	5.	6.

## A conditions for successful solution of the task at hand

Below is a list of *external and internal conditions* that could be important for task solution. Depending on learning tasks these conditions could vary in importance. For example, it could be quite unimportant for the solution of a multiple-choice factual question to consult as many information sources as possible.

**Imagine you would have to actually solve the *present task*<sup>11</sup>. In your opinion, how unimportant or important are those conditions below?**

	very unimportant							very important
	1	2	3	4	5	6	7	
1 Interest in the topic of genetics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 Motivation to solve the task	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 Ability to draw independent conclusions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 Prior knowledge concerning genetics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5 Deep understanding of genetics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6 Knowledge about learning strategies	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7 Knowledge about tasks and task demands	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8 Access to multiple information-sources (e.g., internet)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9 Potential help from others	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10 Sufficient time	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

<sup>11</sup> As indicated on the cover page of the COPES-questionnaire, this questionnaire was printed in a duplex format with the tasks always presented on the left page and the questions pertaining to a specific task always printed on the right page. Thus, the same questionnaire was filled in six times for tasks from all six *Bloom-Categories*. To shorten this appendix, the COPES-questionnaire (A9) and the corresponding tasks (A10, see next appendix) will be presented separately.







17. Imagine you would have to actually solve the *present task*. All information and learning material would be incorporated in a hypertext learning environment. How much time would you plan realistically for the solution of the task? Please, take into account your current standard of knowledge. For example, if you just had to look up a definition, it wouldn't take very long.

**I would need \_\_\_\_\_ minutes.**

18. Imagine you would have to actually solve the *present task*. All information and learning material would be incorporated in a hypertext learning environment. One page in that learning environment explains one central concept (e.g., the structure of DNA or the method of gel electrophoresis). How many different concepts would you have to understand to solve the present task? For a detailed factual question pertaining to one concept, you probably would just have to understand one concept.

**I would need to understand \_\_\_\_\_ concepts.**





6. Below is a list of six categories that describe different kinds of learning tasks. Indicate the category that the task at hand belongs to in your personal opinion. The selection of only one category is possible.

**I would assign the task at hand to following category:**

<input type="checkbox"/>	REMEMBER: Tasks of this kind mainly require to recall and reproduce information from long term memory. The recall of a definition would be an example of such a task.
<input type="checkbox"/>	UNDERSTAND: Tasks of this kind require more than just a reproduction, they require understanding. Rephrasing information in one's own words, writing summaries, transforming information (e.g., explain a diagram), finding examples or classifying information are typical examples of such tasks.
<input type="checkbox"/>	APPLY: Tasks of this kind require the application of procedures or heuristics (e.g., rule of multiplication). Some of these tasks only require the execution of a known algorithm, others do not give a heuristic. Finding the right procedure is part of those tasks.
<input type="checkbox"/>	ANALYZE: Tasks of this kind require a deep understanding of the interrelationship of all components and the overall structure. Drawing independent conclusions from material, differentiating between important and unimportant information, structuring and organizing information or determining an author's perspective are examples of such tasks.
<input type="checkbox"/>	EVALUATE: Tasks of this kind require an evaluation on the basis of idiosyncratic or predetermined criteria and standards. Finding errors or inconsistencies in processes or information or evaluating processes or products (e.g., evaluate cars on the basis of gasoline consumption) are typical examples of such tasks.
<input type="checkbox"/>	CREATE: Tasks of this kind require the creation of a product of one's own. Planning a new product or predicting consequences of changes are typical examples of such tasks. Not only a real creation but also the new and creative organization of given information falls into this category (e.g., write a critical comment).

## A10 – Tasks Representing the six *Bloom-Categories*<sup>12</sup>

### Task of the category *remember*:

DNA is cut by which procedure:

- Treatment with DNA polymerase.
- Use of restriction enzymes.
- Treatment with strong acids.
- Use of ligases.
- ... none of the above mentioned options.

### Task of the category *understand*:

Which of the following errors concerning the analysis of STR profiles could lead to matching an STR profile to the wrong person?

- Exchange of probes in a laboratory.
- Mixture of probes in a laboratory.
- Computer problems in the DAD (German DNA analysis database).
- Faulty labeling in storage.
- None of these errors would lead to this kind of problem.

### Task of the category *apply*:

The Bucket family has collected STR profile of all members. Only the profile of Tiffany's, Mellissa's and Amanda's father, Steve, is missing. All his daughters share the same biological mother, Karen. Construct Steve's STR profile for all 13 loci in the table. Mark the genotypes where both alleles are definitely known.

STR	Karen	Tiffany	Melissa	Amanda	Steve
D3S1358	15, 18	15, 16	15, 16	16, 18	
vWA	14, 18	14, 17	17, 18	17, 18	
FGA	21, 25	21, 24	21, 22	21, 24	
AMEL	XX	XX	XX	XX	
D8S1179	13, 13	13, 13	13, 15	13, 15	
D21S11	29, 31	28, 31	28, 29	28, 29	
D18S51	16, 17	17, 18	17, 18	16, 18	
D5S818	11, 12	11, 12	11, 11	11, 11	
D13S317	11, 11	8, 11	11, 11	11, 11	
D7S820	10, 13	10, 13	9, 10	9, 13	
D16S539	11, 11	11, 11	11, 11	11, 12	
THO1	9, 9.3	8, 9	9, 9.3	9.3, 9.3	
TPOX	8, 11	8, 11	8, 8	8, 8	
CSF1PO	11, 13	13, 13	11, 12	13, 13	

<sup>12</sup> As indicated by the previous footnote, these six tasks were integrated in the COPES-questionnaire (A9, previous appendix) in the empirical study. To keep the appendix short, they are presented separately.

**Task of the category *analyze*:**

Analyze the whole process of STR analysis in detail. Produce an overview table about the most important steps of the STR analysis. Include not only the steps in the laboratory, but also the evidence collection at the crime scene and the interpretation of the results. For each step, list the most important errors, uncertainties or problems.

**Task of the category *evaluate*:**

DNA from dead bodies, blood samples and other biological tissues can be collected in different amounts and different states of degradation. Does degradation and amount of DNA impact all methods of DNA analysis equally? Evaluate all methods of DNA analysis according to the impact of degradation.

**Task of the category *create*:**

What would happen if a new law would be implemented in Germany that allowed the analysis of coding DNA? Describe shortly the most important consequences.

## B17 – Tasks Representing the six *Bloom-Categories*<sup>13</sup>

### First task of the category *remember* (ice-breaker):

#### How can you define short tandem repeat (STR)?

- An STR is a multiple repetition of the sequence GATA.
- An STR is a locus that can be sequenced rapidly.
- An STR is a self-duplicating motive.
- An STR is a short motive of 2 – 7 bases that is repeated.
- An STR is a motive of 10 – 15 bases that can be sequenced rapidly.

### Second task of the category *remember* (criterion):

#### When analyzing a multi-copy Y-STR-locus ...

- ... multiple alleles can be diagnosed per individual.
- ... the genetic material has to come from different individuals.
- ... only the first motives have to be analyzed per individual.
- ... the result indicates contamination.
- ... only one allele can be diagnosed per individual.

### Third task of the category *evaluate* (criterion):

#### Paternity Testing

Imagine you are a molecular biologist and got a request from an employee of Pro Familia (a German organization involved in counseling with regard to family planning issues).

More specifically, this employee describes that Pro Familia wants to offer counseling with regard to the conflicts surrounding the issue of biological paternity. Thus, this employee wants to understand the biological background of paternity testing. He reports to have heard that the analysis of biological paternity always involves Y-STR analysis. Considering this background, he wants the following questions answered:

- Does it make sense to use Y-STR analysis?
- Or are there other methods?
- What kind of method would you recommend? Why?

It is your task to answer the questions of the Pro Familia employee. Please write down a consistent answer that considers all his questions.

---

<sup>13</sup> Originally, these tasks were presented on separate pages and the students participating in this empirical study got these tasks sequentially: Only if the previous task was finished the next task was given and could be started. However, to keep this appendix short, these five tasks will be presented together.

**Fourth task of the category *understand*:****Example**

In the application example involving the identification of the alleged remains of the czar family (mt.3.2.5) the mtDNA analysis was applied. Explain why the mtDNA analysis is best suited for this task as well as for other tasks with an “historical” perspective.

**Fifth task of the category *remember*:****How do you correctly label an mtDNA sequence?**

- The name of the sequence consists of the letters of the first 10 bases, for example ATTGAAGCTA.
- All differences with regard to the rCRS will be added. This number is given as correct label.
- All differences with regard to the rCRS will be listed with position and kind of difference.
- The matches will be determined for every single base and counted. Resultingly, four numbers will be given, for example one for the matches with regard to adenine (A).
- The labeling had to be done by Prof. Anderson or his co-workers.



WWU Münster

**Educational Psychology**  
**Psychological Institute III**  
**Research group Prof. Dr. R. Bromme**

## Questionnaire

Dear Student,

You had time to learn as much as possible about the topic of genetic fingerprinting from the learning environment. In the following, you will answer some questions to test your comprehension of the material.

Your comprehension will be tested by 15 multiple-choice tasks. Each task asks about a main concept from one or maximally two hypertext nodes.

If you did not read all hypertext nodes, you may not be able to solve some tasks. Therefore, if you don't know the answer to a task, you may choose the alternative "I don't know". Besides this alternative, five further alternatives will be presented. **Only one of these alternatives is correct.** Thus, you just have to make **one cross per task!**

Thank you for your cooperation!

In order to facilitate matching of the two questionnaires, please fill in a personal code in the adequate fields in both questionnaires. Your personal code consists of:

**1st character:** First letter of your mother's first name.

**2nd character:** First letter of your father's first name.

**3rd character:** First letter of your own first name.

**4th character:** First letter of your place of birth.

**5th and 6th character:** DOB (day of birth).

(e.g. birthday: 12th of September → The 5th character is a „1“ and the 6th character a „2“;

e.g. birthday: 02nd of September → The 5th character is a „0“ and the 6th character a „2“)

Personal Code:

1.	2.	3.	4.	5.	6.

**1. In which cases is the mtDNA analysis preferred<sup>14</sup>?**

- For identifying victims of accidents.
- For analyzing saliva probes.
- For analyzing old bone findings.
- For analyzing blood samples from a crime scene.
- For paternity testing.
- I don't know the answer.

**2. How do you correctly label an mtDNA sequence?**

- The name of the sequence consists of the letters of the first 10 bases, for example ATTGAAGCTA.
- All differences with regard to the rCRS will be added. This number is given as correct label.
- All differences with regard to the rCRS will be listed with position and kind of difference.
- The matches will be determined for every single base and counted. Resultingly, four numbers will be given, for example one for the matches with regard to adenine (A).
- The labeling had to be done by Prof. Anderson or his co-workers.

**3. How do you diagnose a match between two mtDNA probes?**

- If both mtDNA probes were isolated from the same type of body cells.
- If the mtDNA sequences from both probes match at all positions.
- If the same amount of mtDNA could be isolated from both mtDNA probes.
- If the mtDNA genome has the same length in both probes.
- If both mtDNA genomes possess a D-loop.
- I don't know the answer.

**4. Why is it often possible to conduct an mtDNA analysis for degraded probes, even after a nuclear DNA analysis fails?**

- The mtDNA analysis uses shorter DNA fragments than any analysis of nuclear DNA.
- A cell contains more copies of mtDNA than nuclear DNA. Therefore, the probability of finding longer fragments is enhanced.
- Nuclear DNA degrades faster than mtDNA because of the different structure.
- The mitochondria are smaller than the cell nucleus. Therefore they are not as exposed to environmental factors.
- The mtDNA analysis uses a more sensitive method for DNA extraction and amplification.
- I don't know the answer.

---

<sup>14</sup> All tasks of the learning outcome test were originally formatted slightly different, e.g. more space between lines and with squares to cross instead of circles. However, to keep this appendix short, this more concise format was chosen.



- 5. What should be considered when amplifying degraded mtDNA probes?**
- You should amplify long fragments to enhance speed in order to prevent further degradation.
  - You should not conduct a PCR because the heating of the DNA probe results in further degradation.
  - You should only amplify short fragments because longer fragments do not exist any more.
  - You should only amplify short fragments because they do not demand high temperature during PCR.
  - You should amplify long fragments because they do not demand high temperature during PCR.
  - I don't know the answer.
- 6. Which kinds of bones are best suited for mtDNA extraction?**
- Cranium.
  - Porous parts of the vertebrae.
  - Porous parts of the pelvis.
  - Femoral and lower leg bones.
  - Bones are hard and thus do not contain cells or mitochondria.
  - I don't know the answer.
- 7. Which statement with regard to HVRBASE is incorrect?**
- The database HVRBASE contains mtDNA sequences from the hypervariable regions I and II.
  - The database HVRBASE is openly accessible via the internet.
  - The database HVRBASE contains mtDNA sequences from many countries.
  - The database HVRBASE can be used for labeling sequences.
  - All statements are correct, thus none of it is incorrect.
  - I don't know the answer.
- 8. Which kinds of differences between persons are investigated during mtDNA analysis?**
- The individual length of the circular mtDNA genome.
  - The individual number of cytosine on the mtDNA genome.
  - The individual sequence of the mtDNA at specific regions.
  - The individual size of the hypervariable regions on the mtDNA.
  - The individual existence or non-existence of the D-loop.
  - I don't know the answer.

**9. What are the differences between nuclear and mitochondrial DNA?**

- Nuclear DNA is inherited paternally, mtDNA maternally.
- Nuclear DNA is specific for one individual, mtDNA not.
- Nuclear DNA has 2 copies per cell, mtDNA only one.
- Nuclear DNA is circular, mtDNA not.
- There are no differences, DNA is DNA.
- I don't know the answer.

**10. Which step is not part of the mtDNA analysis?**

- Determination of the exact sequence at the hypervariable region.
- Visual / microscopic analysis of the original material.
- Determination of a band pattern.
- Extraction and purification of the mtDNA.
- Amplification of the mtDNA via PCR.
- I don't know the answer.

**11. What does Heteroplasmy on the mtDNA indicate?**

- Different composition of the blood plasma.
- Slight variations of the mtDNA in one individual.
- Slight variations of the mtDNA between different individuals.
- Analysis of mtDNA from heterogeneous tissue, for example hair and blood.
- Analysis of mtDNA from different mitochondria.
- I don't know the answer.

**12. What conclusion can be drawn most likely from a match between two mtDNA probes?**

- Both probes come from the same person.
- Both probes come from the same maternal line.
- Both probes come from the same paternal line.
- Both probes come from the same local region.
- Both probes come from twins or triplets.
- I don't know the answer.

**13. How can you enhance the discriminatory power of an mtDNA analysis?**

- By repeatedly sequencing the same region with different probe material.
- The discriminatory power cannot be enhanced due to the small size of mtDNA.
- By sequencing larger regions or the whole mtDNA genome.
- By comparing with the sequence of the nuclear genome.
- None of these alternatives is correct.
- I don't know the answer.

**14. Which difficulty arose when the human remains from near Ekaterienburg were analyzed?**

- Cranial bones were missing, therefore facial reconstructions were impossible.
- DNA was strongly degraded, thus no genetic fingerprint could be made.
- No living relatives of the dead persons could be detected for DNA comparison.
- A heteroplasmy for one person was inconsistent with living relatives.
- The Soviet regime forbade exhumation and DNA analysis by foreign scientists.
- I don't know the answer.

**15. Why is the revised Cambridge Reference Sequence (rCRS) used today instead of the CRS?**

- The CRS was not derived from a human.
- The CRS was 10 base pairs too short.
- The CRS was 10 base pairs too long.
- The CRS contained false bases at 10 positions.
- The CRS contained 10 base pairs of HeLa DNA.
- I don't know the answer.



# „To Calibrate or Not To Calibrate?“ Conditions and Processes of Metacognitive Calibration During Hypermedia Learning

## Deutsche Zusammenfassung der Dissertation

Aus evolutionärer Perspektive betrachtet sind Menschen sehr adaptiv. Dem gegenüber ist es eine offene Frage, ob das auch für alltägliche Lernerfahrungen gilt. Modelle des selbstregulierten Lernens (COPEs-Modell; Kapitel 2.1) postulieren, dass Adaptivität eine Kernkompetenz guter Lerner ist. Nur wenn Lerner sich flexibel an Situationen anpassen, werden ihre Strategien den spezifischen Anforderungen gerecht. Beispielsweise sind oberflächliche Strategien nicht ausreichend für eine komplexe Aufgabe, die aufwendige Argumentation erfordert. Andererseits zeigen empirische Studien, dass Lernprozesse meist nicht „perfekt“ ablaufen, eventuell aufgrund unzureichender Adaptation.

Bisher wurden diese Adaptationsprozesse wenig systematisch untersucht. Daher ist ein zentrales theoretisches Argument der vorliegenden Arbeit, dass sich dazu Methoden aus dem *traditionellen Kalibrierungsparadigma* anbieten (Kapitel 2.2). Durch zwei konzeptuelle Erweiterungen des Konstrukts „Kalibrierung“ wird es möglich, mit diesen Methoden die Übereinstimmung zwischen *selbstregulierten Lernprozessen* und *externen Anforderungen* zu diagnostizieren. Ein weiterer Vorteil dieses Transfers liegt darin, dass Probleme des traditionellen Ansatzes in interessante Forschungsfragen umgewandelt werden.

Als Kontext wurde das selbstregulierte Lernen mit Hypermedien ausgewählt, vor allem weil es sich dabei um eine heutzutage allgegenwärtige Lernform handelt. Basierend auf einer inhaltlichen Analyse wurden zwei wichtige externe Anforderungen herausgestellt: *Aufgabenkomplexität* (Kapitel 2.3) und *Textkomplexität* (Kapitel 2.4). Für die empirischen Studien wurden diese Anforderungen systematisch manipuliert: Es wurden Aufgaben mit sechs Kategorien unterschiedlicher Komplexität entwickelt (Anderson et al., 2001) und ein hierarchischer Hypertext mit drei Ebenen unterschiedlicher Komplexität zum Thema „genetischer Fingerabdruck“ erstellt. Dadurch ist es möglich sowohl die Prozesse metakognitiver Kalibrierung (*processes of metacognitive calibration*) systematisch zu untersuchen als auch deren Auswirkung auf das Lernergebnis („*To calibrate or not to calibrate?*“).

Zusätzlich ist es eine weitere offene Frage, ob alle Lerner gleichermaßen adaptiv sind. Die meisten theoretischen Modelle nehmen an, dass *inhaltliches Vorwissen* (Kapitel 2.5) eine entscheidende Rolle beim Lernen spielt. Daher wurden in den empirischen Studien quasi-experimentelle Gruppen verglichen: Studierende der Geisteswissenschaften und der Biologie. Die Rolle anderer Lernercharakteristika ist weniger klar. Die meisten Theorien bestätigen die wichtige Rolle *epistemologischer Überzeugungen* (über die Natur des Wissens und des Wissenserwerbs; Kapitel 2.6). In den empirischen Studien wurden epistemologischen Überzeugungen einerseits mit Fragebögen erfasst, andererseits wurde eine experimentelle Manipulation implementiert, die es erlaubte experimentelle Gruppen zu vergleichen:

Studierende mit neutraler Einführung und mit epistemologischer Einführung (die zu „sophistischeren“ Überzeugungen führt). Als weitere zentrale Frage können daher die komplexen Interaktionen zwischen diesen Lernercharakteristika und den oben genannten externen Anforderungen untersucht werden (*conditions and processes of metacognitive calibration*).

*Studie I: Kalibrierung an Aufgabenkomplexität in der Planungsphase (Kapitel 3.3)*

Studierende (52 Biologen, 50 Geisteswissenschaftler) mit unterschiedlichen epistemologischen Überzeugungen bekamen sechs Aufgaben unterschiedlicher Komplexität vorgelegt und mussten für jede Aufgabe ihren Lernprozess planen. Beispielsweise beurteilten sie die Wichtigkeit, tief elaborierende Lernstrategien einzusetzen.

*Studie II: Kalibrierung an Aufgabenkomplexität in der Ausführungsphase (Kapitel 3.4)*

Studierende (14 Biologen, 21 Geisteswissenschaftler), die entweder die neutrale oder epistemologische Einführung gelesen hatten, bekamen fünf Aufgaben unterschiedlicher Komplexität vorgelegt und mussten diese mit dem Hypertext „genetischen Fingerabdruck“ lösen. Simultan wurden ihre Gedanken und ihre Navigationsstrategien erfasst. Beispielsweise wurde festgehalten, wie lange sie für die einzelnen Aufgaben benötigten und wie viel sie dabei planten. Ihre Antworten sind Indikatoren für ihr Lernergebnis.

*Studie III: Kalibrierung an Textkomplexität in der Ausführungsphase (Kapitel 3.5)*

Studierende (25 Biologen, 26 Geisteswissenschaftler) mit unterschiedlichen epistemologischen Überzeugungen bekamen die offene Aufgabe so viel wie möglich mit dem Hypertext „genetischer Fingerabdruck“ über das Thema mtDNA Analyse zu lernen. Simultan wurden ihre Navigationsstrategien und ihre Bewertung der einzelnen Hypertextseiten erfasst. Beispielsweise wurde festgehalten, wie lange sie für Hypertextseiten unterschiedlicher Komplexität benötigen und wie sie die Verständlichkeit der Seiten einschätzten. Ihr Lernergebnis ergibt sich aus einem anschließenden Wissenstest.

Ergebnisse in Bezug auf die *Adaptivität* der Lerner sind durchweg positiv. Lerner passen ihren gesamten Lernprozess systematisch an Aufgabenkomplexität an. Beispielsweise planen sie mehr tief elaborierende und weniger oberflächliche Lernstrategien für komplexere Aufgaben (Planungsphase, Studie I), sie nehmen sich mehr Zeit für komplexere Aufgaben und schauen sich mehr Hypertextseiten an (Ausführungsphase, Studie II). Darüber hinaus passen Lerner ihren Lernprozess auch systematisch an Textkomplexität an. Beispielsweise bewerten sie komplexere Hypertextseiten als schwieriger und nehmen sich auch mehr Zeit diese zu lesen (Ausführungsphase, Studie III). Ein Vergleich der Studien ergibt, dass Lerner zunächst die (nicht untersuchten) „harten“ externen Anforderungen beachten, beispielsweise die zur Verfügung stehende Zeit, und dass dann Aufgabenkomplexität als wichtigere externe Anforderung wahrgenommen wird als Textkomplexität. Eine weitere Schlussfolge-

rung ist, dass die angewandte Forschungsmethode aus dem traditionellen Kalibrierungsparadigma, gut geeignet ist, diese Fragen der Adaptivität zu untersuchen.

Ergebnisse in Bezug auf den Einfluss der Lernercharakteristika zeigen ebenfalls interessante Effekte. *Inhaltliches Vorwissen* hatte so gut wie keine Effekte in der Planungsphase (Studie I). Der stärkste Einfluss zeigte sich bei der aktiven Anpassung an Aufgabenkomplexität (Studie II): Studierende der Biologie zeigten durchweg weniger metakognitive Prozesse, waren schneller, passten sich weniger an, und hatten (trotzdem) die besseren Lernergebnisse. Bei der Anpassung an Textkomplexität (Studie III) führte mehr Vorwissen zu einer besseren Bewertung der Verständlichkeit der Hypertextseiten und zu besserem Lernergebnis. *Epistemologische Überzeugungen* zeigten in der Planungsphase (Studie I) einen signifikanten Einfluss: Lerner mit „sophistizierte“ Überzeugungen über die Variabilität von Wissen schätzten alle Indikatoren tiefer Elaboration für alle Aufgaben wichtiger ein. In der Ausführungsphase (Studie II) gab es dagegen vorwiegend Interaktionen: Lerner mit der epistemologischen Einführung passten ihre Zeit und ihre Planungsaktivitäten besser an die Aufgabenkomplexität an. Darüber hinaus schrieben diese Lerner längere Antworten und argumentierten besser, waren aber schlechter darin, korrekte Antworten zu geben. Bei der Anpassung an Textkomplexität (Studie III) gab es ebenfalls Interaktionen: Lerner mit „sophistizierten“ Überzeugungen über die Komplexität und Unsicherheit von Wissen hatten eine selektive Lernstrategie (lasen viele Seiten, aber jede nur kurz), die sich besonders bei komplexen Texten zeigte. Diese Lerner zeigten schwächere Anpassung. Allerdings führte diese Strategie zu besseren Lernergebnissen. Insgesamt betrachtet zeigen diese Ergebnisse, dass unterschiedliche Lernercharakteristika (Vorwissen, epistemologische Überzeugungen) sehr unterschiedliche Effekte haben, was nahelegt, dass die Wirkungsmechanismen sich unterscheiden. Darüber hinaus zeigen sich auch komplexe Wechselwirkungen mit verschiedenen Lernphasen (Planung, Ausführung), mit verschiedenen externen Anforderungen (Aufgabenkomplexität, Textkomplexität), und für verschiedene Typen (beispielsweise für domänen-spezifische und domänen-generelle epistemologische Überzeugungen).

Alle Ergebnisse bisher berichteten Ergebnisse wurden dazu genutzt, theoretische Modelle des selbstregulierten Lernens weiter zu spezifizieren und adäquate Förderungsmaßnahmen für selbstreguliertes Lernen zu diskutieren (beispielsweise die schon implementierte Maßnahme zur Förderung „sophistizierter“ epistemologischer Überzeugungen). Darüber hinaus wurden offen gebliebene Fragen genutzt, zukünftige Forschung vorzuschlagen. Eine zentrale Frage dieser Dissertation beispielsweise konnte nicht abschließend beantwortet werden: Bessere Adaptation der Lernzeit an die Aufgabenkomplexität führte in Studie II zu weniger korrekt beantworteten Fragen (aber zu mehr geschriebenen Wörtern). In Studie III dagegen gab es keine systematischen Beziehungen zwischen Adaptation und Lernergebnis. Es ist also weiterhin offen, ob bessere Adaptation in jedem Fall wünschenswert ist.





## Deutscher Lebenslauf

Ich wurde 1975 in Münster in Deutschland als die erste Tochter von Josefa und Lothar Pieschl geboren. Von 1982 bis 1986 habe ich die Grundschule besucht, zunächst in Bad Westernkotten und dann in Lipperode. Von 1986 bis 1995 habe ich das Ostendorf-Gymnasium Lippstadt besucht und dort 1995 mein Abitur gemacht.

Ich habe 1995 angefangen, an der Westfälischen Wilhelms-Universität in Münster Psychologie zu studieren. Im Jahr 2002 habe ich das Studium mit einem Diplom in Psychologie abgeschlossen (Titel meiner Diplomarbeit: "Die kenn ich, das ist eine *Prunella vulgaris*." Pflanzenbestimmen: Prozessanalyse bei Intermediates"). Während meines Studiums habe ich außerdem als studentische Hilfskraft in einigen Projekten mitgearbeitet. In den Jahren 1998 und 1999 habe ich Praktika bei Neurobehavioral Associates, Clinical and Consulting Psychologists (Edmonton, Alberta, Canada) gemacht und dort bei der Durchführung empirischer Studien geholfen. Seit 1999 habe ich außerdem als studentische Hilfskraft der Arbeitseinheit Pädagogische Psychologie von Prof. Dr. Bromme gearbeitet, beispielsweise in einem Projekt zur Evaluation von Lernsoftware.

Direkt nach meinem Diplom wurde ich in der Arbeitseinheit von Prof. Dr. Bromme als wissenschaftliche Mitarbeiterin angestellt. Zunächst habe ich in einem Projekt zur formativen und summativen Evaluation einer Lernsoftware zur Pflanzenbestimmung (Bestimmen Online) mitgearbeitet, das vom Bundesministerium für Bildung und Forschung (BMBF) unterstützt wurde. Anschließend habe ich in einem Projekt mitgearbeitet, bei dem es um den Einfluss epistemologischer Überzeugungen auf Lernprozesse ging und das von der Deutschen Forschungsgesellschaft (DFG) gefördert wurde. Meine Dissertation ist in dieses Projekt eingebettet. Zwischen den Projekten und im Anschluss an das letzte Projekt habe ich Kurse an der Universität gegeben, beispielsweise zu den Themen selbstreguliertes Lernen, Expertise, und Lernen mit Multimedia. Zusätzlich war ich von 2003 bis 2006 eine Kollegiatin des von der DFG geförderten Virtuellen Graduiertenkollegs (VGK) „Knowledge acquisition and knowledge exchange with new media“.

Zur Zeit bin ich neben meiner Anstellung als wissenschaftliche Mitarbeiterin in der Arbeitseinheit von Prof. Dr. Bromme auch Koordinatorin der Special Interest Group Metakognition (SIG 16) der EARLI (European Association for Research on Learning and Instruction) und aktives Mitglied des European Network of Research on Epistemological Beliefs (gefördert durch die DFG), der AERA (American Educational Research Association) und der CogSci (Cognitive Science Society).



## Curriculum Vitae

I was born in Muenster, Germany, in 1975 as the first daughter of Josefa and Lothar Pieschl. From 1982 till 1986 I attended primary school, first in Bad Westerkotten and later on in Lipperode. From 1986 till 1995 I went to the Ostendorf-Gymnasium in Lippstadt where I graduated in 1995 and thus obtained my higher education entrance qualification.

I studied psychology at the Westfälische Wilhelms-University Münster starting in 1995. In 2002 I graduated with the degree of Diploma in Psychology (title of my thesis: “I know this plant, it’s a *prunella vulgaris*.’ Plant identification: Process analysis with intermediates.”). While studying I also worked as a research assistant in multiple projects. In 1998 and 1999 I interned at Neurobehavioral Associates, clinical and consulting psychologists (Edmonton, Alberta, Canada) and helped conduct empirical studies. Since 1999 I worked as a student research assistant in the educational psychology research group of Prof. Dr. Bromme, for example in projects concentrating on the evaluation of learning software.

Directly after my Diploma I was hired as researcher in the research group of Prof. Dr. Bromme. First, I worked in a project about the formative and summative evaluation of educational software for plant identification (Bestimmen Online), sponsored by the German Ministry of Education (BMBF). Subsequently, I worked in a project investigating the impact of epistemological beliefs on learning processes, sponsored by the German Research Foundation (DFG). My Ph.D. thesis is embedded in this latter project. In between projects as well as after the second project, I taught university courses, for example about self-regulated learning, expertise, and learning with multimedia. In addition, I have been a student of the virtual Ph.D. program “knowledge acquisition and knowledge exchange with new media” of the German Research Foundation (DFG) from 2003 till 2006.

Besides my continued work as researcher in the research group of Prof. Dr. Bromme, I am currently also coordinator of the Special Interest Group Metacognition (SIG 16) of the EARLI (European Association for Research on Learning and Instruction) and an active member of the European Network of Research on Epistemological Beliefs (sponsored by the German Research Foundation, DFG), the AERA (American Educational Research Association) and the CogSci (Cognitive Science Society).