Technische Universität München

TUM School of Social Sciences and Technology

Supporting computer-based learning by investigating process data –

Analysis of psychophysiological measurements and log file data

Kerstin Sabine Huber

Vollständiger Abdruck der von der TUM School of Social Sciences and Technology der Technischen Universität München zur Erlangung einer **Doktorin der Philosophie (Dr. phil.)** genehmigten Dissertation.

Vorsitz: Prof. Dr. Anna Keune

Prüferinnen der Dissertation: Prof. Dr. Maria Bannert

Prof. Dr. Doris Lewalter

Die Dissertation wurde am 28.04.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Social Sciences and Technology am 13.07.2023 angenommen.

Acknowledgments

Zuerst bedanke ich mich bei allen, die mich während meiner Dissertation mit Kreativität, Expertise und Motivation begleitet haben.

Ich möchte meiner Betreuerin Prof. Dr. Maria Bannert für ihre Unterstützung, Motivation und ihr konstruktives Feedback danken. Liebe Maria, ohne dich hätte ich diese Arbeit nicht bewältigen können. Danke, dass du mir die Freiheit gegeben hast, meine eigenen Forschungsinteressen zu entwickeln und dass du mich stets ermutigst hast, meine Ziele zu verfolgen, um eine bessere Forscherin zu werden. Außerdem möchte ich meinen Dank an Prof. Dr. Doris Lewalter richten. Liebe Doris, du hast es mir ermöglicht, Teil eines hochkompetenten Teams zu werden, was meine Dissertation entscheidend geprägt und vorangetrieben hat. Die ertragreiche Zusammenarbeit mit dir und dem Team der Toolbox Lehrerbildung hat in erheblichem Maße dazu beigetragen, meine Forschungsarbeit erfolgreich umzusetzen, zu schärfen und zu verfeinern. Des Weiteren hat sich Prof. Dr. Doris Lewalter freundlicherweise dazu bereiterklärt, der Prüfungskommission als zweite Prüferin beizutreten. Ebenso bedanke ich mich herzlich bei Prof. Dr. Anna Keune, die eingewilligt hat, den Vorsitz der Prüfungskommission zu übernehmen.

Ein großes Dankeschön geht an meine Kolleg:innen Lyn, Sabrina, Begüm, Elena, Laura und Steffen für die ertragreichen Gespräche. Euer hilfreicher Input hat mir dabei geholfen, meine Forschung weiterzuentwickeln und damit meine Dissertation voranzubringen. Vielen Dank an meinen Hiwi Stefan für seine verlässliche und wertvolle Hilfe.

Zudem möchte ich Jana, Anna-Theresa und Julia für ihre Unterstützung, die schönen Gespräche und das Writing Retreat danken. Danke, dass ihr meine Artikel korrekturgelesen habt und mir stets hilfreiches Feedback, Motivation und Mut gegeben habt.

Ebenso möchte ich meinen langjährigen Freunden Eli, Joe, Laura P., Laura W. und vor allem meiner Lieblingsschwester Verena danken. Die langen Gespräche mit euch haben mich immer wieder ermutigt und mir stets gute Laune beschert. Eure emotionale Unterstützung hat mir sehr geholfen, durchzuhalten und meine Arbeit voranzubringen. Verena, ich danke dir für dein offenes Ohr, die vielen, langen Telefonate und deine

Hilfsbereitschaft meine Arbeit strategisch zu bewältigen. Ein großes Dankeschön geht auch an meine Eltern, für ihre stetige Unterstützung, Neugier und Motivation.

An dieser Stelle möchte ich meine tiefste Dankbarkeit an meinen Partner Steffen ausdrücken, der mir immer zugehört und geholfen hat Probleme erfolgreich zu lösen. Du gabst mir die emotionale und motivationale Unterstützung, die ich brauchte, um diese Arbeit fertigzustellen. Danke, dass du meine Artikel unzählige Male gelesen und korrigiert hast, egal wie spät es war. Vielen Dank für deine stets kreativen Ideen und hilfreichen Tipps und ich danke dir, dass du mich immer wieder ermutigt hast am Ball zu bleiben.

Abstract

The overall goal of the present research is to investigate the value of process data to understand learning processes and, thus, to promote computer-based learning. The findings of this thesis help educators identify learning processes in computer-based learning environments (CBLEs) aiming to provide meaningful, individualized support to learners (e.g., prompts; see chapters 2 and 7). The present dissertation includes two empirical studies investigating the manifestation of learning processes in process data. Both studies demonstrated how learning processes can be monitored and evaluated while learning with CBLEs.

The goal of Study 1 was to validate psychophysiological measures to identify academic emotions and investigate their impact on learning outcomes. Electrodermal activity (EDA) and heart rate (HR) were measured during computer-based learning to monitor learners' emotional states. Therefore, EDA, HR, and self-report data were gathered from 32 participants in a laboratory setting. To determine the manifestation of academic emotions in EDA, HR, and self-reports, negative emotions were induced using negative connotated learning materials about animal welfare. Participants reported their emotional states directly before and after learning, which were then collated with EDA and HR curves. A significant relationship was found between increased negative emotions and increased EDA and HR. Additionally, EDA turned out to be a significant indicator for learning performance. Furthermore, an explorative analysis revealed that boredom manifested in decreased HR.

Study 2 investigated the impact of navigation behavior on learning performance using log file data and process mining analyses. Therefore, log files and self-report data were evaluated from 58 university students who used a CBLE for two weeks. The results showed a significant increase in learning with a very high effect size. A cluster analysis revealed two distinct learner groups, which differed significantly in their navigation behavior and learning outcomes. Here, the interactivity and the time spent on learning-relevant pages were meaningful indicators for learning outcomes, especially recall and transfer performance. In conclusion, the findings showed that beneficial and detrimental learning processes could be inferred from navigation behavior. Thus, the findings demonstrated that navigation behavior impacts learning outcomes. Moreover, Study 2 presented a feasible approach to monitor and evaluate the interactivity and duration spent in a CBLE, which can be used to promote successful learning.

In conclusion, this thesis demonstrates the importance of process data in investigating and supporting computer-based learning. The two empirical studies shed light on process data from different perspectives to obtain a comprehensive picture of learning processes. The findings contribute to identifying and evaluating learning processes in real time. Moreover, this dissertation presents important theoretical, methodological, and practical implications for further research and theory development.

Zusammenfassung

Das übergeordnete Ziel der vorliegenden Arbeit besteht darin, den Mehrwert von Prozessdaten für ein besseres Verständnis von Lernprozessen und für die Förderung des computerbasierten Lernens zu untersuchen. Die Ergebnisse dieser Dissertation helfen Lehrenden, Lernprozesse in computerbasierten Lernumgebungen (CBLEs) zu identifizieren, mit dem Ziel Lernende sinnvoll und individuell unterstützen zu können (z.B. Prompts, siehe Kapitel 2 und 7). Die vorliegende Arbeit umfasst zwei empirische Studien, in denen die Manifestation von Lernprozessen in Prozessdaten untersucht wurde. In beiden Studien konnte gezeigt werden, wie Lernprozesse beim Lernen mit CBLEs überwacht und evaluiert werden können.

Das Ziel von Studie 1 war die Validierung psychophysiologischer Messungen zur Identifizierung akademischer Emotionen und die Untersuchung ihrer Auswirkungen auf Lernleistungen. Die elektrodermale Aktivität (EDA) und die Herzfrequenz (HR) wurden während des computerbasierten Lernens gemessen, um den emotionalen Zustand der Lernenden zu überwachen. Zu diesem Zweck wurden EDA-, HR- und Selbstauskünfte von 32 Teilnehmenden in einer Laborumgebung erhoben. Um die Manifestation von akademischen Emotionen in EDA, HR und Selbstauskünften zu ermitteln, wurden negative Emotionen durch negativ konnotiertes Lernmaterial über den Tierschutz induziert. Direkt vor und nach dem Lernen berichteten die Teilnehmenden über ihre emotionalen Befindlichkeiten, die dann mit EDA- und HR-Kurven abgeglichen wurde. Dabei wurde ein signifikanter Zusammenhang zwischen erhöhten negativen Emotionen und erhöhter EDA und HR festgestellt. Außerdem erwies sich die EDA als signifikanter Indikator für die Lernleistung. Darüber hinaus ergab eine explorative Analyse, dass sich Langeweile in einer verringerten HR manifestiert.

Studie 2 untersuchte die Auswirkungen des Navigationsverhaltens auf die Lernleistung anhand von Logfile-Daten und Process-Mining-Analysen. Dazu wurden Logfiles und Selbstauskünfte von 58 Universitätsstudierenden ausgewertet, die zwei Wochen lang ein CBLE nutzten. Die Ergebnisse zeigten einen signifikanten Lernzuwachs mit einer sehr hohen Effektgröße. Eine Clusteranalyse ergab zwei verschiedene Gruppen von Lernenden, die sich in ihrem Navigationsverhalten und ihren Lernleistungen signifikant unterschieden. Dabei waren die Interaktivität und die auf lernrelevanten Seiten verbrachte Zeit aussagekräftige Indikatoren für den Lernerfolg, insbesondere für

Erinnerungs- und Transferleistungen. Zusammenfassend zeigten die Ergebnisse, dass aus dem Navigationsverhalten auf förderliche und hinderliche Lernprozesse geschlossen werden kann. Die Resultate demonstrieren somit, dass das Navigationsverhalten einen Einfluss auf die Lernleistung hat. Darüber hinaus wurde in Studie 2 ein praktikabler Ansatz zur Überwachung und Evaluation der Interaktivität und Verweildauer in einer CBLE vorgestellt, der zur Verbesserung des Lernerfolgs herangezogen werden kann.

Zusammenfassend zeigt diese Dissertation die Bedeutsamkeit von Prozessdaten für die Untersuchung und Förderung des computergestützten Lernens. Die beiden empirischen Studien beleuchten Prozessdaten aus unterschiedlichen Perspektiven, um ein umfassendes Bild von Lernprozessen zu erhalten. Die Ergebnisse tragen dazu bei, Lernprozesse in Echtzeit zu identifizieren und zu evaluieren. Darüber hinaus werden in dieser Dissertation wichtige theoretische, methodologische und praktische Implikationen für die weitere Forschung und Theorieentwicklung aufgezeigt.

Included Publications

Parts of this dissertation are two manuscripts, which have been published in two different international peer-reviewed journals. The author of this dissertation is the first author of both articles and played a leading role in the development, conceptualization, data collection, writing, statistical data analysis, and publication-based presentation of these journal manuscripts. The supervisor, Prof. Dr. Maria Bannert, guided the development of both studies with critical reviews, contributed to manuscript revision, and approved the submitted manuscripts.

Manuscript A was submitted to the open-access, peer-reviewed *Journal of Computers in Education* in June 2021 and was accepted for publication in March 2022 (see Appendix A).

Huber, K., & Bannert, M. (2022). What happens to your body during learning with computer-based environments? Exploring negative academic emotions using psychophysiological measurements. *Journal of Computers in Education*. https://doi.org/10.1007/s40692-022-00228-w

Manuscript B was submitted to the open-access, peer-reviewed *Journal of Computing in Higher Education* in July 2022 and was accepted for publication in April 2023 (see Appendix B).

Huber K., & Bannert, M. (2023). Investigating learning processes through analysis of navigation behavior using log files. *Journal of Computing in Higher Education*. https://doi.org/10.1007/s12528-023-09372-3 (published online on 27.04.2023)

Contents

List of Figures

List of Abbreviations

AEQ	Academic Emotions Questionnaire
Ag/AgCl	Silver chloride electrode
CBLE(s)	Computer-Based Learning Environment(s)
CK	Content Knowledge
EDA	Electrodermal Activity
EES-D	Epistemically-Related Emotion Scales (German version)
EK	Evaluation Knowledge
EL501	Disposable snap electrodes for electrocardiogram
EL507	Dry disposable snap electrodes for electrodermal activity
HR	Heart Rate
HRV	Heart Rate Variability
MOOC(s)	Massive Open Online Course(s)
PA	Positive Affect
PANAS	Positive and Negative Affect Schedule
PCK	Pedagogical Content Knowledge
PK	Pedagogical Knowledge
PNS	Parasympathetic Nervous System
NA	Negative Affect
RS-13	Resilience Scale (short version)
SRL	Self-Regulated Learning
SS2LB	Fully-shielded cable assembly for snap electrodes
SS57L	Lead set for electrodermal activity
SNS	Sympathetic Nervous System
TCEK	Technological Content Evaluation Knowledge
TCK	Technological Content Knowledge
TK	Technological Knowledge
TPACK	Technological Pedagogical Content Knowledge
TPEK	Technological Pedagogical Evaluation Knowledge
TPEACK	Technological Pedagogical Evaluation and Content Knowledge
TPK	Technological Pedagogical Knowledge
TTE	Toolbox TeacherEducation (German: Toolbox Lehrerbildung)

1 Introduction

Undoubtedly, computer-based learning environments (CBLEs) gained importance in educational institutions and have been focused in educational research, especially during the COVID-19 pandemic (Eickelmann & Gerick, 2020; Hong et al., 2021; Hoss et al., 2021; S. G. Huber & Helm, 2020). This exceptional situation has demonstrated that computer-based learning is challenging for educators and learners alike (Grewenig et al., 2021; Hoss et al., 2021). Additionally, frontal teaching and traditional lectures have recently been questioned due to their static and uniform setting (Goedhart et al., 2019; Mingorance Estrada et al., 2019). Learning with CBLEs can remedy this issue because learners can study at any time, at their own pace and preferred learning style (e.g., passively watching videos or actively reading and absolving tasks; Adewoye & Olaseni, 2022; Wauters et al., 2010). Moreover, CBLEs offer an interactive, flexible, and unique learning opportunity and provide multiple learning representations (e.g., videos, figures, visualizations; Adewoye & Olaseni, 2022; Goedhart et al., 2019; Patel, 2013).

Nevertheless, promoting computer-based learning can be demanding for educators due to the missing face-to-face communication. In classroom settings, educators can easily monitor, evaluate, and regulate learning processes. For example, if the learner is confused, the educator can intervene immediately and offer support. However, this personal interaction cannot always be provided in CBLEs (Arguel et al., 2017). Moreover, learners need external support and internal regulatory skills to use CBLEs successfully (Van der Kleij et al., 2015). Therefore, educators must rely on methods and approaches to measure learning processes from a distance (Adewoye & Olaseni, 2022; Arguel et al., 2017; Goldman, 2009; Van der Kleij et al., 2015). However, computer-based learning has become indispensable and hosts a wide range of benefits for educators as well as learners. Hence, the question arises of how to support computer-based learning by investigating learning processes.

Research shows that self-regulating learning (SRL) processes (e.g., organizing, monitoring, evaluating) improve computer-based learning (Azevedo & Gašević, 2019; Hattie, 2017; T. McLaughlin & Yan, 2017; Paans et al., 2020; Schneider et al., 2021). However, learners do not always spontaneously monitor, evaluate, and regulate their learning processes, which can hinder computer-based learning (Azevedo et al., 2013).

Process data is a promising approach to measuring and fostering these SRL strategies (e.g., Bannert et al., 2015; Bannert & Reimann, 2012; Duffy & Azevedo, 2015; Engelmann & Bannert, 2021; Pieger & Bannert, 2018; van Alten et al., 2020). Process data include psychophysiological measures, log files, or eye-tracking (e.g., Bannert et al., 2014; Fan et al., 2022; Lim et al., 2021; Malmberg, Haataja, et al., 2019). Furthermore, process data provide information not only about the current state but also about the unfolding learning process (Bannert et al., 2014). By evaluating, for example, psychophysiological data or log files, educators can monitor and regulate the learning process to promote learning performance based on patterns in the data (e.g., Järvelä et al., 2023; Lim et al., 2023). Therefore, process data can be used to provide support from a distance. In this way, educators can compensate for the lack of face-to-face interactions. However, evaluating process data has not yet found its way into everyday learning and teaching (Schneider et al., 2021; Wohlfart & Wagner, 2022). Here lays the foundation of my research goals. I want to address the lack of evaluating computer-based learning using process data. Many studies used a collaborative learning setting to examine the relation between learning processes and physiological measures (e.g., Dindar et al., 2019; Järvelä et al., 2021, 2023) or a laboratory setting to analyze log files (e.g., Fan et al., 2022; Lim et al., 2023; Sonnenberg & Bannert, 2015). However, since computer-based learning mostly takes place at home in a natural setting, I want to focus on self-studying. Moreover, I want to investigate to what extent process data can be used to understand computer-based learning better. The core of this thesis are psychophysiological measurements (i.e., electrodermal activity [EDA] and heart rate [HR]) and log file data, which are introduced in the following paragraphs and described in chapter 3 in great detail.

Psychophysiological data measures physiological responses accompanied by psychological processes (e.g., sweaty fingers when thinking about the next exam). Therefore, it is possible to infer psychological processes from physiological response patterns (Cacioppo et al., 2016; Cacioppo & Tassinary, 1990; Goetz et al., 2022; Pinel & Pauli, 2012). Furthermore, psychophysiological measures have been used extensively in prior and current educational research (e.g., Donker et al., 2018; Eteläpelto et al., 2018; Järvelä et al., 2023; Malmberg, Haataja, et al., 2019; Malmberg, Järvelä, et al., 2019). These circumstances underline the contribution of psychophysiological measurements in education. As a second example of process data, log files are used in the current thesis.

Log files record the interaction with CBLEs (e.g., mouse clicks, duration of stay), which provides information about the learner's navigation behavior. Moreover, log file data can be used to identify patterns in navigation behavior to assess learning processes (Thompson & Markauskaite, 2014). Previous research has shown that certain patterns in navigation behavior are particularly beneficial for learning (e.g., Bannert, 2006; Bannert et al., 2015; Lim et al., 2021). Hence, navigation behavior is a meaningful indicator for learning success.

To investigate psychophysiological measures and log file data, two studies were conducted. Study 1 aims to understand the physiological appearance of academic emotions and their impact on learning outcomes. In addition, Study 1 aims to validate psychophysiological measurements further to identify learning processes, as they are still unattended (Goetz et al., 2022; Loderer et al., 2020; Pekrun & Stephens, 2012). Study 1 showed that EDA and HR were fruitful, given the emergence and progression of emotions and task appraisal during computer-based learning. Furthermore, EDA turned out to be a valid indicator for learning success; for HR, no significant relation could be found. Therefore, HR as a measure for valence in CBLEs requires further research.

In Study 2, log file data were used to examine navigation behavior and its effect on learning performance. The goal of Study 2 was to find patterns in navigation behavior to detect detrimental and beneficial learning processes and their relation to learning performance. Moreover, Study 2 showed a feasible approach to successfully implement a CBLE in a flipped classroom and an algorithm to automatically evaluate log files. Results demonstrated that the duration and interaction with the CBLE were meaningful indicators for successful learning. By marking learning-relevant pages in a CBLE and tracking the intensity of interactions, as well as the time learners spent in the CBLE, instructors can ensure high learning outcomes. Since these factors can be measured and evaluated quite rapidly, the implementation in everyday education is, indeed, feasible.

In conclusion, both studies investigated process data to identify learning processes through two different data modes (psychophysiological data and log files) and data channels (EDA, HR, duration, interactions). As a result, my approach demonstrates how psychophysiological and log file data can be complementary to investigate computer-based learning. Therefore, my findings can be used to advance the research on multimodal data. Furthermore, this work aims to contribute to the research about generating and providing individualized support to enhance computer-based learning.

2 Promoting computer-based learning

CBLEs are used increasingly in everyday life and are nowadays an essential tool for teaching and learning (Ahlan et al., 2014; Arnold et al., 2015; Eickelmann & Gerick, 2020). In the first quartal of 2020, 70% of learners over the age of 16 in Germany used digital learning materials (Statistisches Bundesamt, 2020). Additionally, more and more reviews and meta-analyses are conducted which investigate different factors that impact computer-based learning aiming to promote successful learning with CBLEs (e.g., SRL, Guo, 2022; emotions, Loderer et al., 2020; computer self-efficacy, Moos & Azevedo, 2009; learning outcome, van Alten et al., 2019; Van der Kleij et al., 2015). These facts illustrate the necessity and potential of CBLEs in educational settings. Nevertheless, educators and learners must adapt to teaching and learning with CBLEs since it requires computer and media literacy (Eickelmann & Gerick, 2020) as well as SRL skills (Veenman, 2016). Moreover, the success of computer-based learning needs to be ensured (Adewoye & Olaseni, 2022). An example of measuring learning success is using tasks embedded in the CBLE. Additional feedback after completing the task can benefit learning (Van der Kleij et al., 2015). However, constantly processing feedback requires mental effort and can disrupt the learning process (Wauters et al., 2010). Hence, measuring the learning process implicitly would be a promising opportunity to improve learning.

A well-researched approach for enhancing learning processes is prompts. Prompts can trigger lower-level (e.g., organization, elaboration) and higher-level strategies (e.g., goal-setting, monitoring, evaluation) and can be adapted based on the learners' need for support (Bannert, 2009). Here, the main focus lies on measuring and fostering SRL activities (see Bannert et al., 2015; Bannert & Reimann, 2012; Duffy & Azevedo, 2015; Engelmann & Bannert, 2021; Pieger & Bannert, 2018). SRL describes a goal-oriented learning process composed of cognitive, affective, metacognitive, and motivational activities. In conclusion, depending on individual needs, prompts can trigger cognitive (e.g., organization), affective (e.g., emotional states), metacognitive (e.g., evaluation), and motivational (e.g., self-efficacy) learning activities based on the learners needs (Azevedo et al., 2017).

A study from Lim and colleagues (2023) demonstrated how prompts can be personalized and implemented in a CBLE. Here, process data were used to identify patterns in SRL activities during learning. Based on these patterns, personalized prompts were generated automatically, which enhanced SRL activities. Moreover, a meta-analysis from Guo (2022) demonstrated that the adaptability of prompts is a significant moderator for SRL. Thus, specific prompts (i.e., adapted to the task or situation) improve learning (Guo, 2022). In sum, prompts must be tailored to the individual learner and task to improve SRL and learning outcomes. In order to capture the learners' response to the prompt in real time, evaluating process data is a fruitful approach (Hadwin et al., 2007). In conclusion, process data can facilitate the real-time monitoring and evaluation of computer-based learning without unnecessarily interrupting the learning process (Arguel et al., 2017).

This work wants to contribute to investigating computer-based learning through process data. Therefore, two studies were conducted. In the first Study, psychophysiological data were used to identify academic emotions, which influence learning outcomes (Pekrun & Stephens, 2012). In Study 2, log file data were examined to detect patterns in navigation behavior. In both studies, connections to learning outcomes were made. Both approaches aimed to determine detrimental and beneficial learning processes to enable providing individually tailored support leading to enhanced learning. A detailed description of process data and CBLEs can be seen in chapter 3.

2.1 Challenges and benefits of computer-based learning

Based on Mayer's (2014) *cognitive theory of multimedia learning*, CBLEs epitomize an optimal medium to deliver learning content. Visual and auditory information can be presented simultaneously (e.g., video tutorials), the limited capacity of information processing can be considered (e.g., pausing or repeating a video sequence), and active learning can be supported (e.g., learning tasks including learning relevant feedback and interactivity).

Additionally, computer-based learning holds a host of benefits: learners can study at any time, at their own pace, and according to their learning preferences (J. E. McLaughlin et al., 2014; Wauters et al., 2010). Moreover, CBLEs make knowledge dissemination and examination (e.g., quizzes, tests) faster and easier, resulting in a high degree of freedom and flexibility for educators and learners (Ahlan et al., 2014; Cheok et al., 2017). Furthermore, adaptive support can be provided instantly and regularly through CBLEs, which leads to higher learning outcomes (Van der Kleij et al., 2015). For example, learners can communicate through chat messages with educators or in forum posts with other learners. Also, feedback embedded in tasks can solve problems or reinforce learning behavior (see learning tasks on the platform www.toolbox.edu.tum.de). Instant feedback highlights learners' strengths and deficits, indicating how to achieve their learning goals (Adewoye & Olaseni, 2022; Patel, 2013). These possibilities demonstrate the advantages of CBLEs.

Nevertheless, computer-based learning also poses challenges for educators and learners. Adewoye and Olaseni (2022) propose learner characteristics that impact computer-based learning: attitude towards CBLEs, self-discipline, computer literacy, and motivation. Learners with a positive attitude towards computer-based learning, great self-discipline, high computer literacy, and motivation have significantly more academic success. Moreover, access to computers and the availability of technology are essential factors when it comes to computer-based learning (Adewoye & Olaseni, 2022). Hence, these characteristics must be supported to ensure successful learning. Furthermore, learners need to know how to regulate their learning strategies (van der Graaf et al., 2022). Thus, crucial components of successful learning and reaching goals are continuous elaboration, monitoring, and self-regulation, which are associated with SRL and need to be practiced (Schunk & Greene, 2018).

Nonetheless, not only do learners need to be aware of their own learning processes, but also educators need to monitor and evaluate the learners' behavior (Grewenig et al., 2021). During lectures, it is possible to identify detrimental or beneficial learning processes and solve problems immediately by discussing them face-to-face (Adewoye & Olaseni, 2022). Thus, meaningful indicators must be found to supervise the learning process in CBLEs from a distance. In conclusion, computer-based learning is an auspicious method for teaching and learning. Still, it needs particular know-how from both educators and learners to benefit from it (for a review, see Moos & Azevedo, 2009).

To facilitate the implementation of CBLEs for educators, the *Technological Pedagogical and Content Knowledge* (TPACK) model from Koehler and Mishra (2009) was developed (see Figure 1). It addresses how technology (e.g., CBLEs) can be used meaningfully, purposefully, and profitably to meet specific learning goals. Therefore, three knowledge domains (i.e., content, pedagogical, and technological) and, in particular, their interplay (i.e., pedagogical content, technological pedagogical, and technological content) are presented. Educators should master these knowledge domains when planning media-supported classes (see Figure 1). Moreover, the TPACK model represents an important guideline for planning media-based teaching.

The most meaningful competency for educators using CBLEs is the technological knowledge (TK) and its interactions with pedagogical (TPK) and content knowledge (TCK). The TK describes the knowledge about technologies or digital media. It also includes the knowledge about selecting and using a medium (e.g., CBLE) suitable for reaching learning and content-related goals. Here, TPK refers to which technology or medium is best suited to achieve pedagogical learning goals. TCK addresses the issue of which the technology or medium is most appropriate to reach content-related learning goals (Koehler & Mishra, 2009).

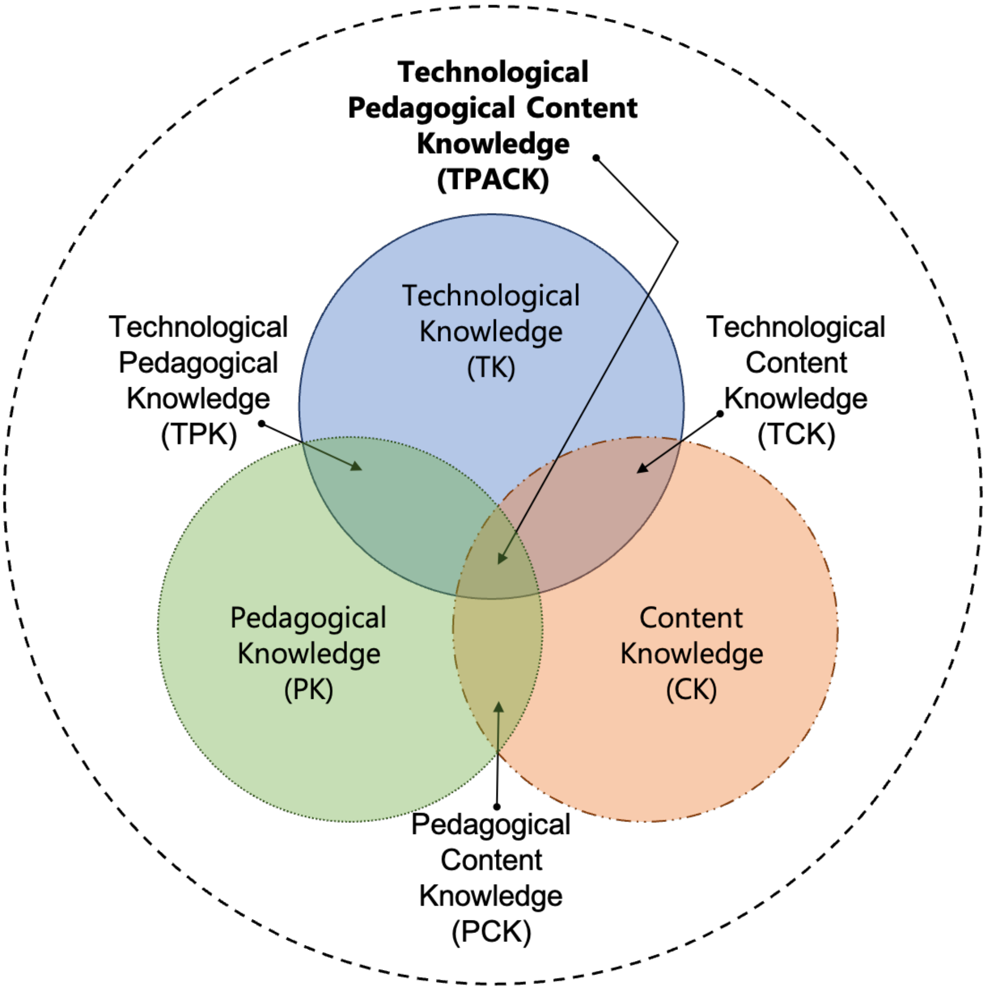


Figure 1. The Technological Pedagogical Content Knowledge Model (TPACK) from Koehler and Mishra (2009, p. 63).

Although this model is very comprehensive, no competency that deals with the knowledge of evaluating and monitoring learning processes is described. Yet, the knowledge about how to evaluate computer-based learning is a crucial component of using a CBLE meaningfully (Ahlan et al., 2014; Azevedo & Gašević, 2019; Eickelmann & Gerick, 2020; Hattie, 2017). Hence, I enriched the TPACK model based on my findings. The adapted model and its implications can be seen in chapter 7.2.

The following chapters present two approaches to investigating and indicating learning processes in CBLEs through process data: measuring academic emotions through psychophysiological measurements and analyzing navigation behavior using log files.

3 Investigating computer-based learning using process data

The most used method to investigate learning success is questionnaires due to the uncomplicated implementation in CBLEs. However, questionnaires measure the current state of knowledge before the learning process has started and after the learning process has been completed (Pekrun, 2020). Nevertheless, measuring the learning process in real time provides information about when and why the learner faced difficulties in the CBLE (Järvelä et al., 2021; Malmberg, Haataja, et al., 2019), which allows individualized and purposeful problem-solving.

A popular and well-researched approach to remotely measuring learning processes is the analysis of process data (e.g., Bannert et al., 2014; Cerezo et al., 2020; Dolak, 2019; Reimann et al., 2014; Schoor & Bannert, 2012; Sypsas & Kalles, 2022). Process data make computer-based learning apparent and observable in real time (Thompson & Markauskaite, 2014).

In addition to the descriptive information (e.g., number of clicks, visit duration, current heart rate, or sweat glands activity), this work explores patterns in psychophysiological data and navigation behavior to identify beneficial and detrimental learning processes. Based on this identification, educators can assess the learning process and provide individual support based on the resulting pattern. Furthermore, developers of CBLEs can implement technologies that evaluate the learning process and present adequate solutions automatically (e.g., Wauters et al., 2010).

A concise presentation of both psychophysiological measurements (see chapter 3.1) and log file data (see chapter 3.2) is presented in the sections below. A detailed description can be found in Manuscript A for psychophysiological measurements (Study 1) and in Manuscript B for log file data (Study 2, see appendices).

3.1 Academic emotions and psychophysiological measurements

What is meant by the term emotion is widely known and frequently used in the general population. However, research does not agree on a specific definition of emotions (e.g., Izard, 2010; Reisenzein, 2007; Schmidt-Atzert, 2009). Nevertheless, there is consensus that emotions are processes that unfold over time and include different components (i.e., cognitive, affective, motivational, physiological, and expressive, Moors, 2009; Scherer et al., 2001). Moreover, emotions can be seen as a response to significant internal or external stimuli enabling a quick reaction to the current situation (Goetz et al., 2022). In order to narrow the concept of emotions, this thesis focuses on the physiological component, which describes the physical concomitants of emotions (e.g., sweaty hands or increasing heartbeat when being excited, Goetz et al., 2022). Additionally, we concentrate on the widely used two-dimensional model, which characterizes emotions in their valence (positive or negative) and arousal (also activation; activating or deactivating, Levenson et al., 2016; Pekrun et al., 2011; Pekrun & Stephens, 2012). Enjoyment and anger, for example, are both activating but differ in their valence. In contrast, enjoyment and relief are positive emotions but differ in their arousal (see Pekrun & Stephens, 2012).

Since this thesis investigates learning-relevant emotions, I address a specific set of emotions. Pekrun and Stephens (2012) introduced the concept of *academic emotions*, which occur in educational settings and are bound to success, failure, learning, and achievement (e.g., anxiety of failing or the joy of passing the exam, Pekrun & Stephens, 2012).

Another theoretical fundament is the *Dual Processing Self-Regulating Model* from Boekaerts (2011), where the importance of emotions during learning becomes apparent. Boekaerts (2011) states that depending on the learners' emotional assessment of a task, learners take the well-being or the growth pathway. Learners who experience negative emotions triggered by, for example, a non-solvable task take the well-being pathway, which prevents knowledge increase. Whereas learners in a positive emotional state (e.g., triggered by a solvable task) take the growth pathway, resulting in a knowledge increase. Furthermore, learners can switch from the well-being to the growth pathway and vice versa (Boekaerts, 2011). Identifying these pathways and detecting the switches in real time is an essential subject of this thesis. In doing so, learners can be individually supported and guided to the growth pathway, resulting in successful learning.

Additionally, the systematic review and meta-analysis from Loderer and colleagues (2020) demonstrate that the impact of emotions on computer-based learning has been researched increasingly over the past years. The authors conclude "that emotions are important drivers of learning in technology-based settings and that learners' emotional experiences can be shaped by the characteristics of those settings" (Loderer et al., 2020, p. 13).

Because (test or exam) anxiety is mainly researched (Pekrun et al., 2002, 2010), little is known about other academic emotions, especially in the context of computer-based learning (Linnenbrink-Garcia & Pekrun, 2011; Pekrun & Stephens, 2012). Therefore, Pekrun and colleagues (2014) suggest focusing on other academic emotions, for example, frustration or confusion.

To address this lack of research, we investigated various academic emotions. First, we explored emotion measurement methods. Since emotions are an internal experience, it is reasonable to merely ask the learners about them using questionnaires (i.e., self-reports, Goetz et al., 2022). Self-reports have been used for a long time, are well-elaborated, and are essential to evaluate academic emotions (Pekrun, 2020). However, self-reports entail notable drawbacks because they are prone to biases (Goetz et al., 2022). For example, learners may exaggerate or understate an emotion to achieve a particular goal (e.g., better grading, Pekrun, 2020). Moreover, only consciously experienced emotions can be reported, which is relevant because academic emotions are not necessarily perceived intensely during learning but significantly influence learning processes and outcomes (Arguel et al., 2017; Loderer et al., 2020). Hence, including objective measurements for emotions can mitigate these problems (Goetz et al., 2022).

Psychophysiological measurements are a promising approach to measuring emotions objectively, implicitly, and in real time (Eteläpelto et al., 2018; Järvelä et al., 2021; Winne & Perry, 2000). Psychophysiology combines psychology and physiology, meaning that psychological processes express in physiological measures. For example, thinking about upcoming exams triggers arousing thoughts, resulting in sweaty hands and a faster heartbeat. Therefore, physiological responses can reveal psychological processes (Cacioppo et al., 2016; Pinel & Pauli, 2012; Potter & Bolls, 2012). Popular psychophysiological measures are EDA (electrodermal activity) and HR (heart rate), which are proven sufficiently to provide insights into emotional states (Boucsein, 2012; A. Lang, 2014; Levenson et al., 2016; Potter & Bolls, 2012). EDA and HR are non-invasive, easy to measure, and sensitive to emotional processes (Berntson et al., 2017; Dawson et al., 2016). I used EDA and HR to measure the learners' emotional states based on the two-dimensional model of emotions. According to psychophysiological research, EDA captures the activation and HR, the valence dimension of emotions (Kreibig, 2010;

P. J. Lang et al., 1993; Palomba et al., 1997). Additionally, I included self-reports to collate the psychophysiological responses to the learners' subjectively experienced emotions. A detailed description of EDA, HR, and the apparatus of emotion measurements can be seen in chapters 5.1.2 and 5.1.3, as well as in Manuscript A.

In conclusion, emotions are crucial for computer-based learning because they influence learning processes and outcomes. Therefore, it is necessary to consider emotions, especially in CBLEs, since there is no face-to-face communication, where emotional states can be determined easily. Using psychophysiological measures during the learning process enables the educator and learner to identify the emotional state in real time and guide the learner onto the growth pathway to improve learning.

3.2 Navigation behavior and log file data

The second approach to investigating learning processes through process data is analyzing log files. Collecting log file data is a simple-to-implement and efficient method to trace the learners' interactions with CBLEs (Cerezo et al., 2020; L. Huang & Lajoie, 2021; Matcha et al., 2019). Log files include data about actions within a CBLE (e.g., mouse clicks, time spent on pages), which characterize the learners' navigation behavior (e.g., systematic: following the structure of a CBLE or explorative: clicking on specifically selected hyperlinks). Furthermore, log files present objective, automated, real-time data revealing dynamic cognitive processes (Schoor & Bannert, 2012) and individual learning behaviors in CBLEs (Azevedo et al., 2013; Malmberg et al., 2010). Although log files are not as fine-grained as think-aloud data or interviews, log files do not depend on learners' recalls or perceptions and can hardly be adjusted by learners (Siadaty et al., 2016). Therefore, analyzing the navigation behavior using log files presents a promising approach to assess learning processes and support computer-based learning.

In addition to analyzing the descriptive event data from log files (e.g., number of clicks, duration of page visits), it is meaningful to examine patterns in navigation behavior (Arguel et al., 2017; Thompson & Markauskaite, 2014). Based on these patterns, insights into various psychological processes can be gained (Thompson & Markauskaite, 2014) and thus may indicate beneficial or detrimental learning. Furthermore, another purpose of identifying patterns in navigation behavior is to predict learning outcomes. It then becomes possible for educators to provide adequate support to learners and to promote learning (Azevedo & Gašević, 2019; Paans et al., 2020).

Process mining is a popular approach to detect patterns in log file data and analyze learning behavior (Dolak, 2019; Lim et al., 2021; Sonnenberg & Bannert, 2016, 2019). Here, log file data is used to create a process model, which illustrates the sequence and flows of the interactions with the CBLE (Bannert et al., 2014; Dolak, 2019). Moreover, the process model can visualize navigation behaviors of different learner groups (e.g., high learning outcome vs. low learning outcome) and learning strategies (L. Huang & Lajoie, 2021; Matcha et al., 2019). A detailed description of how learning processes can be measured using log files and process mining can be seen in chapters 5.2.2, 5.2.5, and Manuscript B in the appendix.

In sum, I want to analyze how log files and navigation behavior can contribute to promoting computer-based learning. Furthermore, I consider the feasibility of evaluating process data in CBLEs.

4 The present research

Since CBLEs do not provide face-to-face interactions as known in traditional classroom settings, it is difficult for educators to monitor individual learning processes and success. Integrating questionnaires for this purpose is reasonable because of their uncomplicated implementation, but they solely measure the current state (e.g., emotions, learning outcomes) before and after learning. Thus, analyzing process data is an effective approach for supervising learning processes in real time. Moreover, indicators for learning success must be identified to provide individualized support. Additionally, finding relations between patterns in process data and learning performance allows improvement of adaptive support. Furthermore, investigating patterns in process data (besides the descriptive analysis of frequencies) is very revealing for identifying and characterizing different learner groups (Bannert et al., 2014).

Two empirical studies were realized to identify indicators for learning success and, therefore, predict learning outcomes. Both studies highlight process data from different points of view to get comprehensive insights into learning processes. Moreover, different patterns for learner groups were detected, and beneficial and detrimental learning processes were identified combined with the questionnaire data.

In the first study (Manuscript A "What happens to your body during learning with computer-based environments? Exploring negative academic emotions using psychophysiological measurements"), psychophysiological measurements (i.e., EDA and HR) were used to assess academic emotions. The study aimed to comprehend the physiological appearance of academic emotions and their impact on computer-based learning. Moreover, Study 1 focused on validating psychophysiological measurements to investigate computer-based learning and learning processes. This exploratory approach was especially fruitful since it has received little attention in prior research.

I structured the research questions and hypotheses in a top-down design, with the broad research question at the top and three following hypotheses. The derived exploratory research question and specific hypotheses are as follows:

RQ: Can psychophysiological measurements provide deeper insights into learning processes?

H1: Negative activating academic emotions cause HR deceleration over time.

H2: Negative activating academic emotions cause increasing EDA over time.

H3: Depending on the learning performance (high vs. low), overall HR and EDA differ.

The second study (Manuscript B: "Investigating Learning Processes Through Analysis of Navigation Behavior Using Log Files") focused on how educators can evaluate and monitor the learning process and progress in CBLEs through navigation behavior. Furthermore, the results from Study 2 aimed to support educators and developers of CBLEs. Educators can monitor the learners' interactions with the CBLE, identify their preferred learning style, and provide immediate support to mitigate the missing face-to-face interaction as known from traditional classroom settings.

To explore log file data and to what extent navigation behavior indicates learning success, I formulated the following research questions and hypotheses:

RQ1: To what extent can navigation behavior predict learning outcomes?

H1: Navigation behavior affects learning outcomes.

H2: Navigation behavior reflects the difficulty level of the learning process.

RQ2: To what extent do learners differ based on navigation and learning behaviors?

H3: Learners with high learning outcomes display different patterns of navigation than learners with low learning outcomes.

Both empirical studies aimed to promote computer-based learning and support learners, educators, and developers of CBLEs. It can be mentioned that Study 1 has helped to establish the previously little-used psychophysiological measurements in the research on computer-based learning. Moreover, Study 2 demonstrated how a CBLE could be successfully implemented in educational settings. Furthermore, Study 2 demonstrated that indicators for learning success and SRL strategies can be found even in an openly available CLBE, where a diverse learner group can navigate freely. In summary, both studies contributed to research on multimodal data. Based on the significant results, it can be inferred how process data can measure learning processes in CBLEs and how support needs to be adapted to promote learning. Both studies aimed to understand learning processes during computer-based learning through multimodal data and to support computer-based learning. Furthermore, my findings show that psychophysiological and log file data can complement each other to investigate computer-based learning. The analyses used to address the hypotheses and research questions can be seen in chapter 5.1.6 for Study 1 and in chapter 5.2.5 for Study 2. Hypotheses and research questions are answered in chapter 7.1.1 for Study 1 and in chapter 7.1.2 for Study 2.

5 Methodology

5.1 Explorative laboratory study (Study 1)

A laboratory study is characterized by a controlled environment, allowing the elimination of confounding variables, resulting in high internal validity. However, the transferability to the real world is often questioned (i.e., low external validity). Psychophysiological measurements, for example, need specific apparatuses and are sensitive to movement artifacts. Nevertheless, these challenges can be handled easily in a laboratory (Potter & Bolls, 2012; Sedlmeier & Renkewitz, 2018).

Indeed, new technologies allow the measurement of psychophysiological variables in the field (e.g., wearables). Nevertheless, compromises must be made regarding handling mobile electrodes and measurement accuracy (Jennings & Allen, 2016). EDA, for example, can be measured on different body parts (see van Dooren et al., 2012). In a laboratory, the most responsive areas (i.e., hand and feet) can be chosen, whereas in field settings, mainly the wrist is used, which fails to show responses when fingers or feet did (Dawson et al., 2016; Payne et al., 2016). Equally, HR can be affected by body movements, which can also be controlled in a laboratory (Potter & Bolls, 2012).

Besides, the lack of research on the relationship between academic emotions, psychophysiological measurements, and CBLEs led to the decision to develop a straightforward exploratory laboratory study. The setup consisted of a screen in front of the participant, including a keyboard, mouse, speakers (left and right of the screen), and the recording unit for EDA and HR measurements (i.e., *BIOPAC MP36*). A camera was installed on the desk, capturing EDA electrodes attached to the fingers. This recording was used to monitor the fit of the electrodes as well as hand or body movements. All artifacts caused by movement (e.g., wiggling fingers, sneezing, leaning back) were noted and removed when processing the data.

Furthermore, the *BIOPAC MP36* was placed outside the participant's field of view because the cables and apparatuses can seem intimidating (Potter & Bolls, 2012). Moreover, items unrelated to the study were removed from the desk and walls to direct the full attention to the learning materials.

5.1.1 Participants

The participants ($N = 32$; 21 females; $M_{\text{age}} = 27.82$, $SD = 2.45$) were recruited by the web-based online system *ORSEE* (Greiner, 2015). The participation was voluntary and uncompensated. To ensure that the learning materials (i.e., video and article) were perfectly understood, the inclusion criterium was being fluent in German. A detailed sample description can be found in Manuscript A in the appendix.

5.1.2 Self-report measures

The study consisted of different emotion scales in a (pre-)posttest design to measure the emotional state before and after learning. Moreover, results from the self-reports were collated with EDA and HR curves. In addition, a resilience scale was included as the learning materials triggered negative emotions. The materials can evoke a stressful experience, which can be better compensated by more resilient individuals (Tugade & Fredrickson, 2004). As a result, resilience impacts the subjective perception of the emotional experience and needs to be controlled when dealing with emotional stimuli. Furthermore, a self-designed domain knowledge questionnaire was included in the pre- and posttest to gather the learning performance. Since measuring emotional states is a crucial part of this study, a detailed description of the scales can be seen below and in Manuscript A (see appendix).

The pretest included the German version of the Positive and Negative Affect Schedule (PANAS, Krohne et al., 1996), the seven-item German short version of the Epistemically-Related Emotion Scale (EES-D, Pekrun et al., 2017), the short form of a resilience scale (RS-13, Leppert et al., 2008), one multiple-choice question about the political opinion on animal husbandry in Germany, the current diet of the participant, and a self-developed questionnaire about the prior knowledge regarding the topic to be learned. The posttest comprised PANAS, EES-D, Academic Emotions Questionnaire (AEQ, Pekrun et al., 2011; Titz, 2001), the content-related questionnaire, and questions about age and gender.

The PANAS measures two factors of the emotional experience: Positive Affect (PA) and Negative Affect (NA, Watson et al., 1988). Whereas a high PA is accompanied by concentration and exhilaration, a low PA expresses in sadness and lethargy. In contrast, a high NA reflects irritability and anxiety; a low NA implies balance and calm

(Krohne et al., 1996). The PANAS scale includes 20 adjectives, 10 for PA (e.g., active, excited, proud) and 10 for NA (e.g., upset, scared, ashamed) on a 5-point Likert-scale with the response alternatives “very slightly or not at all”, “a little”, “moderately”, “quite a bit” or “extremely” (Krohne et al., 1996; Watson et al., 1988). Additionally, the short version of the EES-D was included because this scale focuses on learning-related emotions. The seven-item EES-D asks for the intensity of the items surprised, curious, excited, confused, anxious, frustrated, and bored on a 5-point Likert-scale with the possible responses “not at all”, “very little”, “moderate”, “strong”, or “very strong” (Pekrun et al., 2017). Based on Pekrun and colleagues (2017), these emotions often occur during computer-based learning. Moreover, the PANAS and EES-D cover the most prominent dimensions of emotions (i.e., activation and valence) and can therefore highlight the relation between EDA (activation) and HR (valence). For example, a high EDA should come along with a high value for active or scared. In combination with HR, statements about the valence of emotions can be made (see chapter 3.1).

Besides activation and valence, I wanted to examine the emotional perception of the learning situation. Therefore, I included a subscale of the AEQ. Since the AEQ collects information about the emotional experience of the previous learning situation, it was included solely in the posttest. The AEQ scale consists of class-, exam- and learning-related scales, which can be used separately. Each item relates either to experiences before, during, or after learning on a 5-point Likert-scale from “strongly disagree” to “strongly agree”. The AEQ scale includes eight subscales regarding enjoyment, hope, pride, anger, anxiety, shame, hopelessness, and boredom (Pekrun et al., 2011; Titz, 2001). Because the AEQ comprises 232 items, we used the 45 items scale, which refers to emotions during learning, to avoid overwhelming the participants.

Regarding learning performance, I designed an 11 items questionnaire based on the content of the learning material. Pre- and posttest included identical items to gather knowledge increase or decrease, resulting in a measure for learning performance. The questionnaire included 10 multiple-choice items with three to four response alternatives (e.g., “Was bedeutet animal-turn?”, “Welche Vorschrift(en) haben andere Länder bezüglich der Tierhaltung?“) and one open question (i.e., “Wie wird die Tierschutznutztierhaltungsverordnung abgekürzt?”). In order to minimize guessing, the possibility of answering “I don’t know“ was included in every question.

5.1.3 Psychophysiological measures

To collect psychophysiological data, I used the *BIOPAC MP36* system. EDA was measured using the SS57L lead set, including EL507 disposable snap Ag/AgCl pre-gelled electrodes. According to the standards of recording EDA (see Boucsein, 2012; Dawson et al., 2016), two EL507 electrodes were placed on the palmar phalanges of the index and middle finger of the non-dominant hand (see Figure 2). Before applying, the skin was prepped with saline to remove excessive oil (see Potter & Bolls, 2012). This step ensures that the electrodes adhere well and that electrical contact with the skin is provided (see Boucsein et al., 2012). Due to the electrodes' size and the lead set's pinch, either the proximal (see C and D in Figure 2) or the medial (see A and B in Figure 2) phalanges were chosen, depending on the provided area on the fingers. Since the highest density of eccrine sweat glands was found on the palms in general, both proximal and medial phalanges are very responsive to emotional stimuli (Dawson et al., 2016; van Dooren et al., 2012).

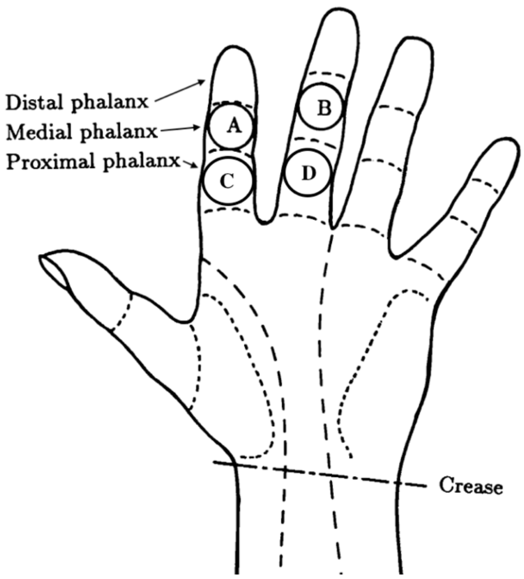


Figure 2. Placement of Electrodes on Medial (A, B) and Proximal (C, D) Palmar Phalanges (based on Boucsein, 2012, p.105).

Furthermore, HR was collected using the fully shielded SS2LB cable and disposable snap Ag/AgCl pre-gelled EL501 electrodes. The HR measurement standards were followed to attach the electrodes (see Berntson et al., 2017; Potter & Bolls, 2012). According to the lead-II placement and the Einthoven triangle, three electrodes were placed on the upper body: two electrodes under the collarbone (left and right) and one on the left side of the ribcage (see Figure 3). Before attaching the electrodes, the skin was wiped with isopropyl alcohol to remove dead skin cells and oil, minimizing electrical impedance (see Potter & Bolls, 2012).

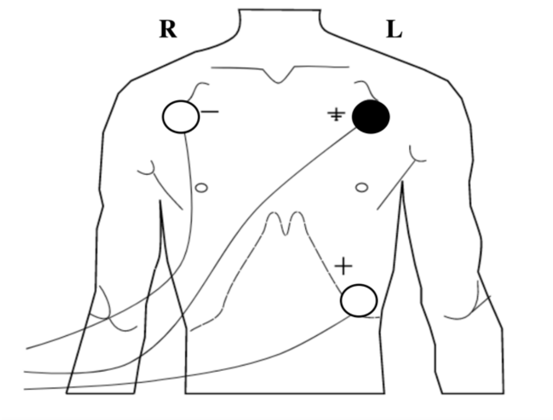


Figure 3[1]**.** Lead-II Placement of Heart Rate Electrodes (based on Fortin-Cote et al., 2019, p.2).

To record and process EDA and HR data, I used the software *Biopac Student Lab* 4.1 (BIOPAC Systems Inc., 2019), sampled with 1kHz (meaning 1000 samples per second). A high sample rate guarantees high-quality and valid data after smoothing the curves and removing artifacts (Boucsein et al., 2012). Furthermore, the research software *iMotions* v8.1 (iMotions, 2019) was used to navigate the study (e.g., providing the learning materials automatically on the participants' screen) and observe the participants' activity (e.g., showing the participant's screen).

5.1.4 Learning environment

The learning environment, consisting of a 6-minute video and a scientific article about conventional pig farming, was selected to induce negative emotions and to lead the participants on the well-being pathway (see chapter 3.1). The video was a report made by a public broadcaster. It contained recordings of conventional pig farms made by animal welfarists (i.e., ARIWA, ANIMALS' ANGELS) and scenes of political discourses about animal welfare. The report began with dramatic music and a voiceover about the illegally

[1]The electrode placement in Figure 3 in Manuscript A differs from the presentation provided here, which is neither methodological or technical issue, nor effected the results.

recorded pigsties and their dreadful condition. The goal of the animal welfarists was to expose animal welfare scandals. This video segment triggered scare and dismay. After the scenes from the pigsties, a political event was shown, where the federal minister of Food and Agriculture (Germany) demanded that this kind of recording should be banned. Moreover, she affirmed that violations of animal welfare are punished.

Subsequently, facts are presented which give evidence that pig farms and the welfare of animals are controlled only sporadically. Moreover, public veterinarians were interviewed with the result that after reporting violations, the veterinarian was withdrawn from controlling the pigsties. These scenes demonstrated the imbalance between reality and politics, which evoked anger. Then the illegally recorded scenes from pigsties continued, leading to sadness and distress. Continuously, the animal protection act and the political discussion were reported auditorily. Meanwhile, an employee is shown who killed piglets, which activated distress and anger. In the end, the reporter concluded that the violation of the animal protection act was not punished sufficiently, which triggered frustration. Altogether, the report induced severe negative emotions.

Afterward, the participants had to read a scientific article about conventional pig farming and the animal protection act from Bruhn and Wollenteit (2018). The article includes paragraphs and a detailed description of animal welfare's legal basis, making it difficult to read and understand. Therefore, the negative emotional states of the participants continued and triggered frustration or boredom. The study aimed to make the change in emotions psychophysiological measurable and visible (i.e., changing EDA and HR).

The learning materials were pretested (N = 5) to ensure they activate negative emotions. The pretest results can be seen in the supplementary material of Manuscript A.

5.1.5 Procedure

After arriving in the laboratory, participants were involved in a short talk to get used to the laboratory setup. Afterward, they were assured that the data would be collected using a pseudonym and anonymized after processing. Moreover, the procedure, application of the electrodes, and functionality of EDA and HR measurements were explained in detail. The topic of the study was shared without disclosing hypotheses or research interests to avoid biasing the participants. After every step of the procedure, the instructor confirmed that the participant could ask questions anytime.

Then, the participants filled in the questionnaires (pretest, see chapter 5.1.2). When the participants finished the questionnaires, the electrodes for EDA and HR were applied (see chapter 5.1.3). Before attaching the electrodes, the instructor commented on every step to inform the participants about the procedure of applying the physiological equipment. After all electrodes and cables were installed, a rest period of five minutes followed. The participants were instructed to relax and sit still. During this period, the instructor ensured that the electrodes were applied correctly and that the gel, skin, and electrodes could evenly hydrate. Moreover, this non-stimulated measure served as a baseline (details see chapter 5.1.6).

After the rest period, the video started, followed by the article (see chapter 5.1.4). The participants were instructed to pay undivided attention, understand the content, memorize as much information as possible, and take as much time as needed to read the article. After the participants finished reading, the instructor removed all electrodes and asked the participant about their feeling. Then, the participants completed the posttest (see chapter 5.1.2). Ultimately, the participants were informed about the research interests and questioned their opinions about the learning materials. A detailed illustration of the procedure can be seen in Manuscript A.

5.1.6 Data processing and analyses

All questionnaires were collected using the online survey tool *SoSci Survey* (Leiner, 2019). Psychophysiological data were recorded using the *Biopac Student Lab* (BIOPAC Systems Inc., 2019) and processed using the MATLAB-based application *PhysioDataToolbox* v0.5 (Sjak-Shie, 2019). Self-report and processed psychophysiological data were analyzed using *SPSS Statistics 26* (IBM Corp., 2020) and *JASP* (JASP Team, 2020).

Before processing EDA and HR, a visual inspection was conducted to identify artifacts or measurement errors. If necessary, artifact removal or smoothing routines were performed (see Potter & Bolls, 2012). Afterward, the recommended standard procedures (Boucsein et al., 2012; Potter & Bolls, 2012) and the instructions of the software creators were followed (Sjak-Shie, 2019) to analyze the psychophysiological data. As a result, the baseline was subtracted from the raw EDA data, and RR-intervals for heart rate variability (HRV) analyses were generated (see Manuscript A).

An ANOVA with repeated measurements was conducted to test whether negative activating academic emotions cause HR deceleration (H1) and an increase in EDA over time (H2). Moreover, a trend analysis was performed to examine the linear progression of EDA and HR curves. Additionally, two simple linear regressions were executed with EDA and HR as a predictor and learning performance as the dependent variable. Moreover, a One-Way ANOVA was conducted to investigate whether overall HR and EDA differ depending on the learning performance (H3). A detailed description and illustrations of individual steps of data processing and analyses can be seen in Manuscript A in the appendix. Major results can be found in chapter 6.1 and an in-depth interpretation in chapter 7.1.1, as well as in Manuscript A.

5.2 Field study (Study 2)

Besides the highly controllable laboratory study, a less controllable but more lifelike field study was realized. A field study holds the advantage that it takes place in a natural environment. Therefore, the testing conditions resemble everyday situations, resulting in a high external validity (Sedlmeier & Renkewitz, 2018). Ideally, data from a university seminar could be evaluated without interfering (e.g., including a CBLE that would not usually be used), which is a unique feature of this study. A university seminar was monitored over an entire semester, and navigation data were gathered through log files. The seminar was designed as a flipped classroom with regular self-study phases, where the participants used a CBLE to study a particular topic. After each self-study phase, the domain knowledge was measured via an online questionnaire. A detailed description can be seen in chapter 5.2.4 and Manuscript B (see appendix).

Furthermore, questionnaires about usability, motivation, design, and acceptance were included after learning with the CBLE to get information about how the participants experienced the usage. Since we focus on the data during learning, the user experience is not addressed further.

As an additional characteristic, the log file analysis can be mentioned. Since log files can be collected without the learners being aware, undistorted results can be expected (Sedlmeier & Renkewitz, 2018).

5.2.1 Participants

The sample consisted of 62 teacher-training students, including 41 females, 19 males, one diverse individual, and one without specification (M_{age} = 22.18, *SD* = 2.51). All participants attended a seminar about planning, conducting, and analyzing teaching in the summer term of 2020. Regular attendance, learning with the CBLE, and completing tasks were mandatory to pass the course.

5.2.2 Instruments and measures

The most important measure was the knowledge test (also referred to as content-related questionnaire), which was developed in collaboration with the developers of the CBLE based on its contents. The questionnaires have existed since 2016, and the items were continuously improved with statistical analyses: examining if learning outcomes increased significantly and testing the item wording by calculating the item difficulty index (see Jonkisz et al., 2012). The content-related questionnaire aimed to get information about the knowledge achieved after learning (i.e., posttest), under consideration of prior knowledge (i.e., pretest). Therefore, the quality and efficacy of the learning materials can be assessed and adapted if necessary. The study design included a pre- and posttest, from which a difference value was calculated (i.e., learning performance) to measure knowledge increase.

The knowledge test comprises multiple-choice items organized into three difficulty levels based on *Bloom's Taxonomy* (recall, comprehension, and transfer; Bloom, 1956). The level of *recall* means remembering or recognizing facts. The level of *comprehension* stands for understanding and paraphrasing. The level of *transfer* refers to designing and planning new structures (Churches, 2008). To minimize guessing, the possible response "I don't know" was included in every item. Item examples can be seen in Manuscript B in the appendix. All questionnaires were embedded in the online survey website *unipark.com* (Tivian XI GmbH, 2022) and analyzed using *SPSS 26* (IBM Corp., 2020).

The log files were generated from the software plugin *matomo* (*Matomo*, 2022), which tracks various websites. The log files included, for example, the number of visits, visit duration, user ID, number of actions (i.e., clicks), page URLs, and more. After extracting the log files from matomo, they were revised for process mining analyses.

Here, the software *Disco* was used (Fluxicon BV, 2022). A detailed description of the data analysis procedure can be seen in chapter 5.2.5 and Manuscript B (see appendix).

5.2.3 Learning environment

As CBLE, the German open educational resource *Toolbox TeacherEducation* was used (TTE, Lewalter et al., 2018a). This hyper-multimedia, openly available learning platform combines disciplines involved in teacher education (i.e., educational psychology, subject didactics, and teaching subjects). The TTE provides evidence- and competency-based content to achieve a transfer from theoretical knowledge to teaching-related practice. It was used in several seminars and has been proven to accomplish a significant learning gain (Lewalter et al., 2018b, 2020, 2022; Titze et al., 2021).

The TTE comprises diverse multimedia presentations according to the multimedia principle to optimize learning success (Butcher, 2014). Components are summaries of the current state of research, enriched by figures, visualizations, and video tutorials. Another component is the scripted instructional videos, which emulate scenes from school classes. Here, the focus is to combine the three teacher education disciplines through a practical example. Moreover, learning tasks were developed, which can be used to check the knowledge about a particular topic. After each response, the learner receives extensive feedback to deepen the knowledge or solve problems. Additionally, accompanying didactic material is provided, which contains specific recommendations for teaching structure. Especially for the discipline “teaching subject” (i.e., Fachwissenschaft), dynamic mathematical visualizations were created to represent mathematical principles and correlations. In conclusion, according to individual requirements, the described components can be used separately or in combination during class and self-study phases.

5.2.4 Procedure

The seminar took place in the summer term of 2020 and lasted ten weeks, from May until July. It included five online meetings with the educator and four self-study phases. The first session served as an introduction to the TTE and to clarify organizational issues. Moreover, the participants completed the pretest, which included items about all topics covered in the seminar (see Figure 4). At the beginning of each questionnaire, participants were informed about data protection and had to give their consent by marking a box; otherwise, the questionnaire could not be started.

After the first online meeting, the first self-study phase started, where the participants learned about “Motivational Activation” using the TTE (see Figure 4). Each self-study phase followed an online meeting, where questions and issues about the previously learned topic could be discussed with the educator. Additionally, the participants completed the posttest about the previous topic (detailed seminar schedule see Figure 4). The self-study phases were two or three weeks long, and the tasks were to work with the TTE, read an additional scientific article, and write a summary of the content.

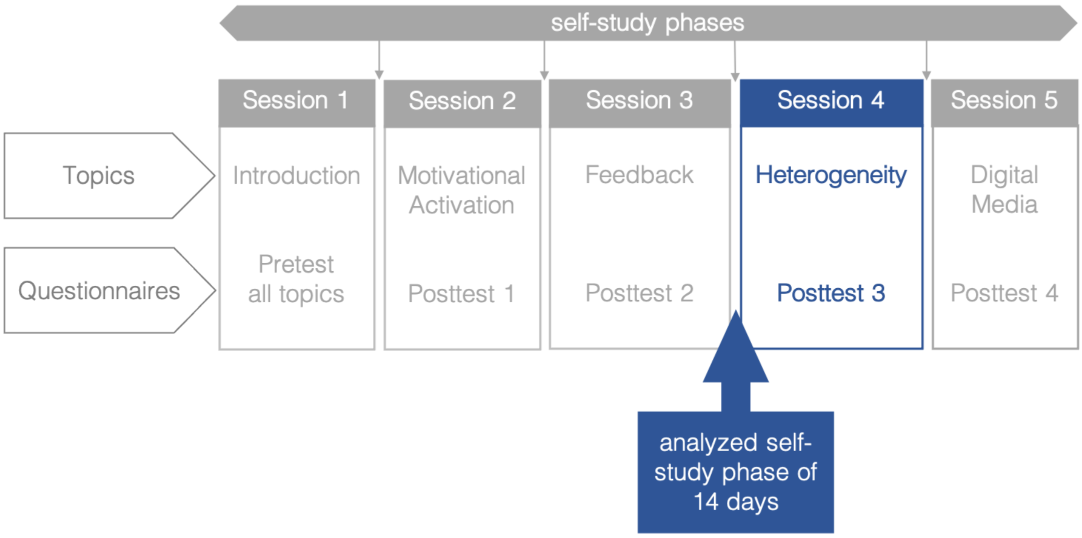


Figure 4. Seminar Schedule and the Analyzed Self-Study Phase Before Session 4.

In Study 2, a specific learning unit was analyzed: one online meeting and one self-study phase (14 days total). The topic for this learning unit was “Heterogeneity” (see Session 4 in Figure 4). This self-study phase was chosen to achieve meaningful results due to the high learning performance (24.8% increase) in combination with the number of available log files (40 out of 58). Although the posttest took place in the online meeting, not every participant completed the questionnaire. Because the questionnaire data must be matched with the respective log file, some data sets could not be used for further analyses. Therefore, the sample size varied but was still satisfactory.

5.2.5 Data processing and analyses

At first, the raw log files were filtered for meaningful variables (i.e., user ID, duration, page title, actions). Afterward, the page titles were labeled, and significant categories were created (i.e., learning-relevant, learning-irrelevant, videos, and orienting). A Python algorithm was developed to aggregate the time spent on particular pages for each visit and in total. Moreover, the Python package pm4py was used to prepare the log file data for process mining analyses. Here, the participants' navigation behavior while learning with the TTE was visualized. In sum, this algorithm evaluated log files automatically and can be adapted and used for future research.

In order to analyze whether the participants achieved a significant learning performance, a paired samples t-Test was conducted. Next, the relations between navigation behavior, learning outcomes, and difficulty level were examined (H1 and H2). Here, I aimed to show that navigation behavior can predict learning outcomes. Additionally, a hierarchical cluster analysis and a One-Way ANOVA were carried out to determine specific navigation patterns for participants with high learning outcomes compared to participants with low learning outcomes (H3). The resulting clusters were used for process mining to analyze to what extent navigation behavior predicted learning outcomes (RQ1) and to what extent participants differ based on navigation and learning behavior (RQ2). A detailed description of data processing and analyses can be seen in Manuscript B (see appendix). Major results can be found in chapter 6.2 and their interpretation in chapter 7.1.2 as well as in Manuscript B.

6 Summary of studies and major results

6.1 Manuscript A (Study 1): "What happens to your body during learning with computer-based environments? Exploring negative academic emotions using psychophysiological measurements"

To support learning with CBLEs, detrimental and beneficial learning processes must be detected. One approach is to investigate emotions during learning (i.e., academic emotions). Research has proven that positive emotions lead to successful learning (Duffy et al., 2020; Loderer et al., 2020). For example, easy and solvable tasks trigger positive emotions, resulting in a knowledge increase (Kang et al., 2008; Pekrun & Stephens, 2012). In contrast, tasks that seem complex or unsolvable activate negative emotions and hinder successful learning (Baker et al., 2010; D'Mello & Graesser, 2014). Therefore, academic emotions can impact learning outcomes (Boekaerts, 2011). Indicators for beneficial or detrimental learning processes can be determined by making academic emotions measurable through psychophysiological measurements.

Academic emotions are usually assessed using self-reports (e.g., Boekaerts, 1999; Eteläpelto et al., 2018; Magno, 2011; Pekrun et al., 2011, 2017), which require a conscious experience of emotions. However, it is comprehensible that learners are not constantly aware of their emotional state. Moreover, self-reports are a subjective and static pre-post measure, which can lead to measurement biases (Laarni et al., 2015; Slater, 2002). Hence, an objective measure can shed light on unconscious emotional states during learning with CBLEs. Furthermore, the shift from beneficial to detrimental learning processes can be captured in real time (Arguel et al., 2017; Boekaerts, 2011). Therefore, I focused on psychophysiological data as an objective real-time measure since they have been proven to capture emotions (Levenson et al., 2016). Because emotions can be defined by two dimensions (i.e., activation and valence), I chose one measure for each dimension: EDA for activation and HR for valence (see Berntson et al., 2017; Dawson et al., 2016). EDA and HR are easy to measure, non-invasive, and well-elaborated (Potter & Bolls, 2012).

The study aimed to explore how academic emotions manifest in EDA and HR. Furthermore, I want to find patterns that indicate beneficial or detrimental learning processes to support learning with CBLEs. Finally, I want to investigate if EDA and HR can be used as an indicator for learning. Therefore, negative, activating emotions were induced utilizing carefully selected learning materials to analyze if the hypothesized patterns in EDA and HR can be found.

My findings revealed that negative emotions significantly increased after learning. Furthermore, increased EDA was a significant indicator for low learning outcomes. For HR, no significant correlation with learning outcomes could be found. Nevertheless, the trend that increasing HR led to low learning performance could be visually observed, leading to the conclusion that high EDA and HR had a negative impact on learning. Moreover, I categorized learners into three distinct groups based on their learning performance: high, middle, and low. These groups exhibited variations in their EDA and HR levels. Specifically, learners with high learning performance showed lower EDA and HR values compared to those with low learning performance.

In summary, I demonstrated that academic emotions manifest in EDA and HR. Moreover, increasing EDA and HR indicated negative, activating emotions in an educational setting. In addition, it can be highlighted that increasing EDA turned out to be a significant indicator for decreasing learning performance. As a result, changes in EDA and HR represent changes in academic emotions.

6.2 Manuscript B (Study 2): "Investigating learning processes through analysis of navigation behavior using log files"

Nowadays, it is undoubtedly, that teaching and learning with CBLEs are essential and valuable in academic life. Especially since traditional lectures are more and more questioned due to their lack of room to address individual needs of learners (Goedhart et al., 2019; Mingorance Estrada et al., 2019). CBLEs encounter these issues due to their interactive character, where learners can pursue their individual learning preferences and needs (Goedhart et al., 2019).

Research about SRL showed that monitoring and regulating learning is crucial when learning with CBLEs. (van Alten et al., 2020, 2021). Consequently, educators and learners must track and evaluate learning with CBLEs to promote successful learning and support individual needs (Arguel et al., 2017; Paans et al., 2020).

Real-time measures (e.g., log files, physiological data, think-aloud, eye-tracking) are a well-elaborated approach to tracking and evaluating learning processes (Fan et al., 2022; K. Huber & Bannert, 2022; Lim et al., 2021; Reimann et al., 2014). Research suggested that patterns in navigation behavior (e.g., captured through log files) can predict learning success (Bannert, 2006; Bannert et al., 2015). However, these methods are still underused in everyday educational life (Schneider et al., 2021).

This study analyzed log files to investigate participants' interactions with a CBLE. My goal was to prove the impact of navigation behavior on computer-based learning and to find patterns in the navigation behavior that indicate beneficial or detrimental learning processes. Moreover, I applied a process mining approach to visualize these patterns. Consequently, I expounded on the value of log files in the context of learning with CBLEs.

Data from a university seminar were gathered, including 58 valid data sets for learning performance and 40 log files, which were matched with the respective learning outcomes. A 14-day self-study phase was evaluated, where the participants learned with a CBLE (i.e., TTE). Learning outcomes were measured through a self-designed knowledge test, categorized into three difficulty levels (recall, comprehension, and transfer; see chapter 5.2.2). Results showed that knowledge increased significantly after working with the TTE for all difficulty levels. The variables duration, actions, and learning-relevant correlated with recall and transfer performance. Therefore, I demonstrated a significant relationship between navigation behavior and learning outcomes. The cluster analysis revealed two significant groups of learners (high and low performers), which differed significantly in their navigation behavior and learning outcomes. High performers spent more time in the TTE and interacted more intensely (i.e., more actions) than the low performers. Low performers showed significantly lower learning outcomes for recall, comprehension, and transfer knowledge. My process mining approach supported these results: high performers reveal a meaningful navigation pattern (i.e., a loop between text and task pages).

In conclusion, I elucidated that log files are a fruitful method to monitor learning processes and provide the opportunity to promote individual needs. Moreover, I demonstrated that navigation behavior can predict learning success. Furthermore, I was able to illustrate different patterns indicating beneficial and detrimental navigation behavior.

7 Discussion and conclusion

Research is moving towards the use of multichannel data for a more “complete” understanding of learning. However, the measurement approaches and learning indicators (and consequently the interpretation) could look different for each data mode and channel (see Azevedo & Gašević, 2019). Therefore, in my work, I focus on two different data modes (i.e., psychophysiological and log file data) and, within each data mode, different data channels (i.e., EDA and HR, duration, and interaction). Here, the goal is to tap into the complementary roles of these measurements of learning processes and how they relate to learning outcomes.

Since openly available CBLEs, also called MOOCs (Massive Open Online Courses), are available for heterogenous learner groups, providing individualized support is demanding yet, required to promote diverse learner preferences (Kovanović et al., 2019). My research contributes to the automatic generation of individualized support (e.g., prompts) based on evaluating psychophysiological measurements and navigation behavior. Moreover, my findings help to extract indicators for timing and adapting support. Previous research (e.g., Kovanović et al., 2019; Wong et al., 2019) addressed how learners handle support in CBLEs, whereas my findings can be used to investigate the emotional reaction to the provided support and how this reaction impacts navigation behavior and learning outcomes.

The present research explored process data for identifying indicators for learning success. Furthermore, patterns in process data were researched to determine beneficial and detrimental learning processes in CBLEs. In this way, educators can understand learning processes in CBLEs and provide adaptive and individualized support. This allows learners to adjust their learning strategies and educators to shape the learners' activities to improve learning.

The empirical studies presented in this thesis introduced two approaches to investigate the manifestation of computer-based learning in process data. Study 1 examined the relationship between emotions and computer-based learning. Therefore, academic emotions were assessed using psychophysiological measurements during learning. It was shown that EDA is an appropriate indicator for learning success. Furthermore, the linkage between HR and learning processes is worthy of further investigation. In Study 2, navigation behavior in CBLEs was measured and evaluated through log files. In addition, a dendrogram identified learner groups, which showed different learning performances and specific patterns in navigation behavior. Thus, navigation behavior could provide information about learning success. These results show that log files offer a contribution to the investigation of computer-based learning.

7.1 Interpretation of central findings

The following sections provide an interpretation of the central findings and answers to both studies' hypotheses and research questions. In addition, I discuss how my research on psychophysiological measures and log file data contributes to the approach of multimodal data in learning.

7.1.1 Psychophysiological measures, academic emotions, and learning performance

Study 1 addressed whether psychophysiological measurements can provide insights into emotional states and learning processes. Therefore, I investigated if psychophysiological measures can be used as indicators for academic emotions to predict learning performance (e.g., Pekrun et al., 2011, 2017; Pekrun & Stephens, 2012). To do so, EDA and HR data were merged (which were recorded during learning) with self-reported emotions (which were measured before and after learning), and different statistical analyses were performed. In the following paragraphs, I provide an in-depth answer to the research question from Study 1, which can be found in chapter 4.

The most meaningful results in my research in Study 1 were that EDA turned out to be a significant predictor for learning performance and that a significant linear relation between EDA and learning performance was found. Here, high learning performance was associated with low EDA, meaning low EDA indicated learning success. However, statistical analyses revealed that HR was not a significant predictor for learning performance. Although, a visual inspection of the HR curve indicated that, as learning performance increased, HR decreased. This means that participants with high learning performance had lower HR compared to participants with low learning performance. These findings indicate that low levels of EDA and HR, therefore, a low activation benefited learning.

In addition, the groups of high, middle, and low learning performance showed significantly different EDA and HR. Therefore, depending on the psychophysiological pattern, learning performance differed, meaning increasing EDA and HR interfere with learning success. As a result, depending on the learning performance, EDA and HR differ, which is in line with H3 (see chapter 4). These patterns show the distinction between the well-being and the growth pathway (see chapter 3.1 and Boekaerts, 2011). Learners on the well-being pathway were characterized by low learning performance and emotionally high activation. This group probably assessed the task more negatively than the group with high learning performance. Therefore, they were more emotionally negative, resulting in low learning performance. Conversely, learners on the growth pathway were characterized by high learning performance and an emotionally low activation due to the positive task assessment. As a result, the well-being pathway can be defined by high EDA and HR, whereas the growth pathway can be associated with low EDA and HR.

7.1.1.1 Interpretation of the results from heart rate analyses

Since HR analyses did not reveal the expected outcomes, a more detailed discussion of these findings is needed. The results showed that HR and self-reported negative activating emotions increased over time, contradicting previous findings about HR and arousing media messages (cf. A. Lang, 2014; A. Lang et al., 2009). Therefore, H1 cannot be supported. The emotionality of the stimulus material (i.e., video and article) could clarify this discrepancy. Since the participants reported significantly higher negative activating emotions, the stimulus material can be considered highly emotionally activating. I assumed that the video would put the participants in an emotionally negative state and that the article would maintain this emotionally negative state. Both of these assumptions occurred. However, I did not expect the negative emotional activation to be at such a high level even while reading. Consequently, the question of the relationship between activation, valence, and HR arises. In order to understand the impact of different stimuli on the HR, a short detour into the nervous system must be made.

In the research about psychophysiology, HR is a common measure for valence. A decreasing HR is a valid indicator of unpleasant stimuli (Ijsselsteijn et al., 2000; P. J. Lang et al., 1993, 1997; Palomba et al., 1997). Yet, due to the complex mechanics of the nervous system, HR can also be sensitive to emotional activation if the activation is very high, which could be replicated (see chapter 6.1 and Manuscript A). The link between activation and valence regarding the HR underlies the dual control of the heart. Its pace is regulated by both autonomic nervous branches, the parasympathetic (PNS) and the sympathetic nervous system (SNS; Berntson et al., 2017; A. Lang et al., 2009; Levenson et al., 2016). Both systems influence how fast the heart beats, depending on which system is activated. The activation of the PNS leads to HR deceleration, which is associated with attention. The activation of the SNS results in HR acceleration, which is related to emotional activation (A. Lang, 2014). Therefore, HR can be a measure of valence but also activation. Nevertheless, since the PNS is faster and more dominant than the SNS, the activation of the SNS must be powerful to overcome the parasympathetic activation (Shaffer & Ginsberg, 2017).

Indeed, my results support this argumentation and demonstrate that the highly emotional stimulus material activated the SNS, and therefore an increase in HR was measured. Hence, the learning environment was too emotionally activating to show that unpleasant stimuli manifest in HR decrease (A. Lang, 2014). In addition, the HRV analysis demonstrated that the SNS controlled 62.9% of HR, the PNS controlled 29.8%, and the remaining 7.23% is related to thermoregulation and can be disregarded. A. Lang and colleagues (2009) stated that further research is needed to investigate the extent to which the PNS or SNS controls the HR. Here, the value of my results becomes evident. In a highly activating, emotionally negative learning environment, I could reveal that the SNS is more dominant than the PNS, resulting in an HR increase instead of the expected HR decrease.

A second explanation could be the task difficulty. Tasks that cause a high cognitive load lead to a higher increase in HR than tasks that elicit a low cognitive load (Cranford et al., 2014). Thus, HR data is a valuable indicator for cognitive load (Bolls et al., 2001; Haapalainen et al., 2010). Also, my results suggest that HR displayed cognitive load and emotional activation rather than the valence of emotions. Hence, adding qualitative data (e.g., an interview after learning) or a control group could encounter this issue.

In conclusion, to what extent HR can be used to indicate valence in the context of learning is questionable. However, identifying the valence of academic emotions is crucial in supporting computer-based learning. Negative emotions need to be solved to enhance learning, whereas positive emotions do not need further support, because they benefit learning (Goetz & Hall, 2013). Nevertheless, my research results are an important step toward using multimodal data. Since HR measurement has received little attention in educational research (Molenaar et al., 2023), it is significant to use my approach for future research. A discussion of the implications of these findings can be found in chapter 7.2.

7.1.1.2 The location of academic emotions in the Dual Processing Self-Regulating Model

First, the trend of self-reported decreased boredom after the learning session was observed, which is in line with the verbal feedback from the participants. They reported great interest in the topic, and thus, boredom decreased. Five participants scored one point higher on the EES-D scale (see chapter 5.1.2) in the posttest, which were defined as "bored". Also, seven participants scored one (n = 5) or two (n = 2) points lower in the posttest on the EES-D scale, which were defined as "not bored". Moreover, bored participants showed a significantly lower HR than less bored participants, suggesting that HR can measure valence but not for activating emotions. Here, the switch between the well-being and the growth pathway can be identified in real time, which enriches the research from Boekaerts (2011). During the video sequence, all participants (i.e., bored and not bored) were on the same pathway, but as soon as the reading part began, the participants went on different pathways. Bored participants took the well-being and not bored participants the growth pathway. This separation happened probably due to the task assessment, which is in line with my findings about psychophysiological patterns and learning performance. Participants, which assessed the task as boring, showed a lower HR which can be associated with lower cognitive load and low emotional activation, which was described in the paragraphs above. However, the sample size for this investigation was relatively small, so the relationship with learning performance is not entirely resolved. Certainly, this attempt is merely exploratory, and more research is needed to support this result.

In conclusion, I demonstrated that negative emotions (e.g., frustrated, distressed, scared, upset) significantly increased after learning, indicating a negative task assessment and a negative emotional state while learning. Therefore, I confirmed that emotions unfold over time (see Scherer et al., 2001), and I showed that academic emotions also evolve in this manner. Concerning the taxonomy of academic emotions from Pekrun (2006), I demonstrated that negatively activating academic emotions are physiologically associated with high EDA (which supports H2) and HR (which is not in line with H1). However, to distinguish individual emotions, self-report data is still necessary. Thus, I contributed to the research that psychophysiological measurements are a promising approach to investigating academic emotions (see Eteläpelto et al., 2018; Järvelä et al., 2021; Winne & Perry, 2000). In addition, EDA and HR increased significantly over time. Hence, increased EDA and HR were indicators of the development of negative emotions. Moreover, my results demonstrated that psychophysiological data is an appropriate tool to measure academic emotions and predict learning outcomes. Thus, I confirmed that psychophysiological responses reveal psychological processes, concluding that EDA and HR provide insights into academic emotions and learning processes. These findings comprehensively answer the RQ for Study 1 (see chapter 4).

7.1.2 Log file data, navigation behavior, and learning success

Study 2 investigated to what extent log file data and navigation behavior can predict learning success (see RQ1 in chapter 4). The results demonstrated that the navigation behavior, defined by duration, actions, time spent on learning-relevant, orienting, video, and learning-irrelevant pages, differs between learners (see RQ2 in chapter 4). Moreover, I wanted to further the research about computer-based learning and how educators and learners can be supported to use CBLEs successfully. Therefore, I addressed the challenges of computer-based learning and investigated how they can be compensated using log file analyses. The results are discussed in detail in the following subsections.

7.1.2.1 The definition of high and low performer

The most significant finding of my research and the answer to RQ2 in Study 2 (see chapter 4) was that the dendrogram revealed two distinct learner types: low and high performers (the dendrogram can be seen in Manuscript B). The results showed that the learning types differed significantly in their learning outcomes and navigation behavior, which aligns with H1 and H3 (see chapter 4). The high performers showed significantly higher recall, comprehension, and transfer performance (which supports H2) but had a similar prior knowledge compared to the low performers. Concerning navigation behavior, the high performers stayed significantly longer on learning-relevant, orienting, and video pages. Therefore, the positive relation between time spent on learning-relevant pages and learning performance underpins prior research (Jeske et al., 2014; Narciss et al., 2007). In addition, the high performers also showed a high duration for learning-irrelevant pages, which is not unusual since they had an overall higher duration of stay in the CBLE. These results support the hypotheses from Study 2 (see chapter 4).

Apart from the above-described duration of stay in the CBLE, the high performers also showed a significantly higher interaction (i.e., number of actions) with text, video, orienting, literature pages, and tasks implemented in the CBLE compared to the low performers. Hence, my findings contribute to determining and characterizing learner types, which allows the prediction of learning success based on measuring interactions and duration (see RQ1). Depending on the identified learner type, support can be tailored to the learners' needs to enhance computer-based learning. It can be assumed that high performers had better SRL skills than low performers. Therefore, presenting regulatory strategies to low performers (e.g., using prompts) can improve their SRL skills and guide them onto the growth pathway. In order to generate individualized support, future research should ask learners about their SRL skills (e.g., at the first login to the CBLE). Moreover, regularly measuring the SRL skills can derive information about the efficacy of the provided support. Thus, the support can be adapted to the skill level (e.g., metacognitive instead of cognitive prompts).

Moreover, my findings demonstrated that passively watching a video also contributed to learning success. Therefore, high performers applied passive (i.e., watching videos) and active (i.e., solving tasks) learning activities, which is also stated by Matcha and colleagues (2019). Furthermore, my findings replicated previous research that the interaction with orienting pages, as an indicator of regulatory activities, is positively associated with learning (Bannert, 2003; Bannert et al., 2014).

In addition, the navigation behavior of the high performers was characterized by a looping pattern between text pages and tasks, which suggests that they read the text, tested their knowledge by completing respective tasks, and then returned to learning (which is in line with H3). This pattern supports the assumption that high performers apply advanced SRL skills and provides an answer to RQ2. The SRL model from Zimmerman and Moylan (2009) describes cyclical phases of SRL: learners begin with setting goals and planning to learn (analyzing the task to acquire knowledge about "Heterogeneity"), then learners carry out the task (learning with the TTE, reading a text), after that, learners evaluate their learning progress using tasks embedded in the TTE. As a result, the looping pattern executed by the high performers illustrated the cyclical SRL phases from Zimmerman and Moylan (2009). Noteworthy is that Study 2 took place in a natural environment which was not designed especially for investigating SRL skills compared to prior studies (e.g., Fan et al., 2022; Lim et al., 2023; Sonnenberg & Bannert, 2015). In fact, the TTE is an open platform and hence, a more open-ended than purpose-built CBLE. This study also shows that meaningful indicators can be extracted even at that level of granularity (i.e., not as fine-grained or complex as other studies) available on such platforms (see also Wong et al., 2019).

As described in chapter 2, SRL can impact learning; thus, measuring learners' SRL skills (e.g., after the initial login) can provide insights into their experience with self-regulation. The outcome of the initial SRL questionnaire could be used as a baseline measure for the individual learner. Therefore, poor interaction and low learning performance can be interpreted more precisely. Thus, support (e.g., prompts) can be adapted to the learners' individual needs. Inexperienced learners do not always have the ability to regulate their learning strategies (Narciss et al., 2007); thus, tailored prompts can help to introduce SRL skills (e.g., triggering lower-level strategies using motivational or cognitive prompts). Otherwise, repeated cognitive prompts could be detrimental to skilled learners since they already have the ability to organize and elaborate their learning strategies (Bannert et al., 2015; Narciss et al., 2007). Furthermore, analyzing SRL skills can support interpreting the navigation behavior because learners with high self-regulation skills show different navigation behavior than less proficient learners (e.g., Bannert et al., 2014, 2015). Hence, prompts considering personal learning preferences and SRL skills can enhance learning (Guo, 2022). Consequently, prompts can be used to guide the learners onto the growth pathway, resulting in high learning outcomes.

Since the CBLE used in Study 2 (i.e., TTE) presents an openly available platform, it can be defined as a MOOC. In a MOOC, a diverse group of learners is represented. Moreover, learners can navigate freely, meaning the given linear structure is not always followed (Wong et al., 2019). Thus, providing support for diverse learner groups and adapting the support to the navigation behavior is challenging. Hence, indicators must be found to determine the timing and type of support that the individual learner requires. For example, a course in the TTE comprises different components (e.g., videos, visualizations, tasks, text), which demand different amounts of time to interact with. Videos, for example, can be skipped, fast-forwarded, or repeated. Therefore, the time learners spend on individual pages is not always meaningful. A promising and upcoming approach to this problem is combining different data modes and channels, also called multimodal or multichannel data (Azevedo & Gašević, 2019; Molenaar et al., 2023). The present thesis contributes to the research on multimodal data by linking log files to psychophysiological data. My findings showed that active interaction and long duration on learning-relevant pages benefited learning. In order to get a more comprehensive insight into the learning process, EDA and HR can be considered. Here, low levels of both channels indicated high learning outcomes. Since the study about EDA and HR took place in a laboratory, the next step would be integrating psychophysiological measurements into a field study on MOCCs. There is a wide range of studies about the relation between psychophysiological measures, especially EDA, and collaborative learning (for a review, see Molenaar et al., 2023). Therefore, I suggest addressing the research gap in investigating learning with MOCCs through psychophysiological measures and log file data in self-studying. As a result, psychophysiological data can be explored in a realistic environment, leading to valid findings and meaningful implications for future research.

7.1.2.2 Navigation behavior and learning outcomes

Besides the distinction of learner types, my findings demonstrated that the overall duration of stay in the CBLE and on learning-relevant pages, as well as the number of actions, had a significant relation with the posttest score (see RQ1 and H1 in chapter 4). Therefore, active interaction and an extended stay in the CBLE contributed to learning success. Hence, learners must invest time, especially on learning-relevant pages, and interact actively with a CBLE to achieve high learning outcomes. Hence, I could replicate the positive relation between time spent on learning-relevant pages and learning outcomes (e.g., Jeske et al., 2014; Narciss et al., 2007). Previous research also showed that absolving learning tasks during computer-based learning improves learning performance (Narciss et al., 2007). Furthermore, I confirmed the finding from Lim and colleagues (2021) that re-reading is an indicator of successful learners.

Since the knowledge test was parted into three different difficulty levels using *Bloom's Taxonomy* (see chapter 5.2.2), the difficulty levels with navigation behavior could be correlated. The results revealed a significant relation between recall, transfer performance, overall duration, actions, and time spent on learning-relevant pages (which supports H2). Thus, I demonstrated that an extended stay, especially on learning-relevant pages, promotes recall and transfer performance. Moreover, this finding contributes to the research from Bannert and colleagues (2015) as well as from Sonnenberg and Bannert (2015), who showed that monitoring the learning process correlated positively with transfer performance.

In conclusion, the answer to my RQ1 is that I demonstrated that the time spent on learning-relevant pages and the associated interactivity (i.e., the number of actions) are essential factors in predicting learning success. Specifically, the extended stay on learning-relevant pages and the active interaction with the CBLE led to high recall and transfer performance. Additionally, I showed that log file data and navigation behavior are suitable tools for gathering information about learning processes. Furthermore, I demonstrated that examining the navigation behavior of learners is effective for predicting learning outcomes. Additionally, I illustrated that beneficial (i.e., high performers) and detrimental (i.e., low performers) learning processes could be visualized through patterns in navigation behavior (which answers RQ2 and supports H3). Therefore, educators can identify beneficial and detrimental learning processes and provide individual support to promote learners.

Moreover, my research showed how to counter the missing face-to-face interaction and presented a method that is easy to implement a CBLE in daily academic life. Therefore, my findings follow the research from Reimann and colleagues (2014), who stated that log files are a meaningful approach for analyzing learning processes. Furthermore, I replicated previous findings that log files provide essential information about individual learning and navigation behavior (Azevedo et al., 2013; Malmberg et al., 2010).

7.2 Methodological and theoretical implications

The present research demonstrates that learning processes manifest in psychophysiological data and navigation behavior. Moreover, the studies validated that EDA and the interaction and duration can predict learning success during computer-based learning. These findings can be used to support educators and learners in evaluating and, thus, promoting learning with CBLEs. Given the lack of theoretical frameworks for evaluating computer-based learning using process data in educational settings, this work wants to fill this gap by drawing methodological and theoretical implications.

From a methodological perspective, Study 1 provides an empirical approach to measuring academic emotions physiologically in a laboratory setting. While EDA was a reliable measure for activating emotions, HR could not represent the valence of emotions, as claimed in H1 (see chapter 4; a detailed discussion of this result can be found in chapter 7.1.1.1). The highly emotionally activating learning materials used in Study 1 have probably been the reason for this issue. Therefore, I recommend considering the emotionality of learning materials and, thus, reconsidering if HR can be an appropriate measure for emotional valence. If the learning material triggers intense emotions (e.g., frustration, distress), based on my research, it is not advisable to use HR as a measure of emotional valence. EDA was, as hypothesized, a reliable measure for activation (see Manuscript A and chapter 7.1.1).

In addition to using EDA solely from an objective, quantitative perspective, I suggest seeing EDA as an indicator for the subjective quality of the learning experience. Therefore, based on the detected emotion, prompts can trigger affective learning activities (Azevedo et al., 2017). Moreover, the emotional response to the prompt can be measured through EDA and HR and, thus, be adapted to shape the emotional learning experience. Additionally, asking the learners about their perceived learning performance can give a conclusive picture of the qualitative experience and also about the design of the CBLE (Loderer et al., 2020).

Moreover, my findings contribute to the SMA grid (Self-regulated learning processes, Multimodal data and Analysis) from Molenaar and colleagues (2023). The SMA grid shows that linking log file data and EDA has received little attention and was primarily investigated in collaborative learning settings (Molenaar et al., 2023). Therefore, my approach demonstrates that navigation behavior can complement EDA data to get insights into the emotional and behavioral processes in CBLEs.

7.2.1 The adaptation of the Dual Processing Self-Regulating Model

At this point, I would like to introduce specific theoretical implications. Based on the findings from Study 1, the *Dual Processing Self-Regulating Model* (Boekaerts, 2011) can be extended (see Figure 5). Therefore, I compartmentalized the well-being and growth pathways in greater detail and added a physiological dimension. In addition, I present a strategy for how learners can exit the well-being and move to the growth pathway based on measuring EDA. Additionally, I segmented each pathway into specific stages, describing the route from the initial task assessment to the final learning performance and the exit strategy from the well-being to the growth pathway. Therefore, each factor impacting learning performance is illustrated (see Figure 5). Furthermore, I implemented an intermediate step on the exit route, where prompts are tailored to the learners' characteristics. This is a crucial step since prior research demonstrated that individualized prompts improve learning (Azevedo & Gašević, 2019; Guo, 2022; Lim et al., 2023). Therefore, individualizing prompts is crucial for guiding the learner onto the growth pathway.

In conclusion, prompts can be generated automatically based on physiological behavior but must be tailored to the individual learner and task. Additionally, SRL skills and learning preferences must be considered. For example, experienced learners do not need detailed cognitive prompts, and inexperienced learners could be overwhelmed with metacognitive learning strategies. Hence, an elaborated collection of prompts must be prepared. It is especially beneficial to learning if prompts and feedback are written personally (Guo, 2022). For example, teachers know the strengths and weaknesses of their students precisely and can therefore ensure ideal individual support. Moreover, advanced learners can adapt the prompts to their needs and link their own resources (e.g., respective literature, YouTube videos, or self-authored documents).

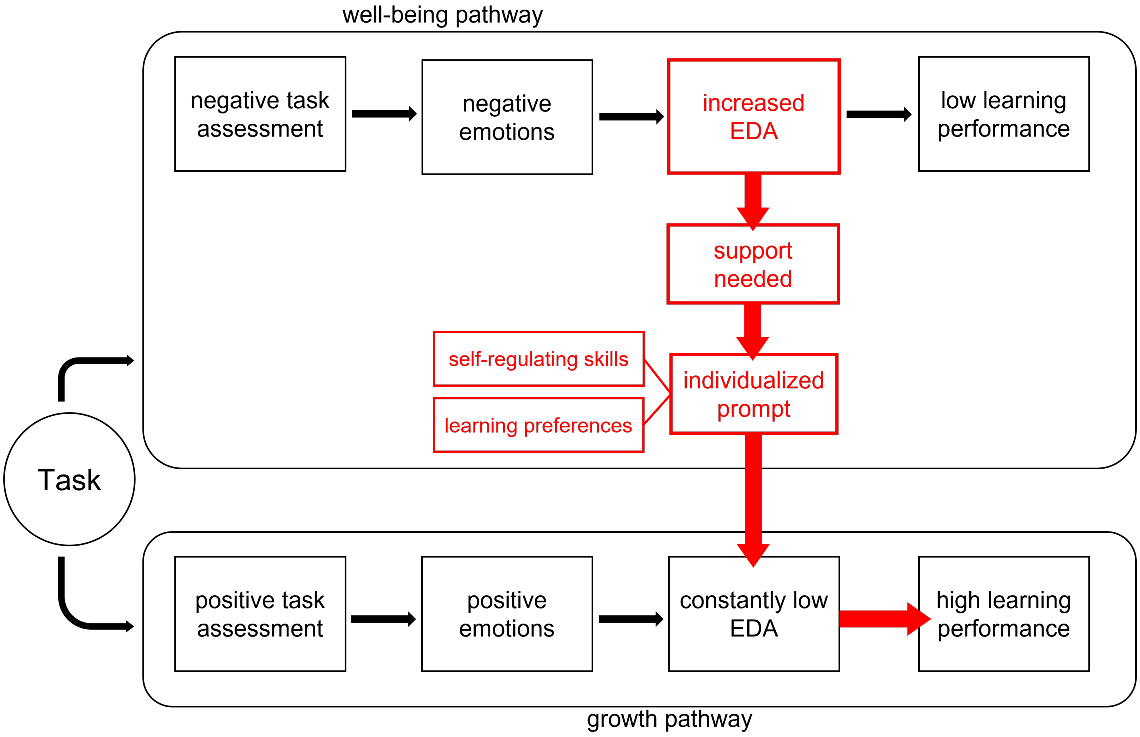


Figure 5. The Adapted Dual Processing Self-Regulated Model (Boekaerts, 2011, p. 410) including Physiological Dimensions and Exit Strategy.

The growth pathway starts with a positive task assessment that triggers positive emotions, as manifested by constantly low EDA, indicating high learning performance (see Figure 5, path on top). The well-being pathway is defined by a negative task assessment, which induces negative emotions, determined by steadily increasing EDA, indicative of low learning performance (see Figure 5, the path below). While Boekaerts' model (2011) solely describes that learners can switch between pathways depending on cues they perceive during learning, my extended model shows how the particular path can be identified in real time. A second unique feature of my model is the additional exit route from well-being to the growth pathway and the introduction of how to follow the exit strategy (see Figure 5, red path). As soon as EDA steadily increases, support must be provided to guide the learner onto the growth pathway. I suggest including an algorithm that automatically generates prompts based on the educator's input for each learner since prompts must be adaptive (tailored to the learner) and task-specific (tailored to the task; Guo, 2022). If the EDA steadily decreases after receiving the prompt, according to my research, previous problems have been solved, and the learner has moved onto the growth pathway.

Furthermore, the continuous measurement of EDA can be used as a “safety net”. Here, my contribution to research on multimodal data becomes evident. For example, if there is no physiological response to the prompt, the navigation behavior must be considered to identify the learners’ state. Additional support is needed if the learner does not interact actively or fails a task after receiving the prompt. A second case example could be that EDA increases shortly after a prompt. Since EDA measures emotional activation (Boucsein et al., 2012), a positive (e.g., excitement) or negative emotion (e.g., frustration) could be triggered through the prompt. Hence, the physiological response is identical (i.e., a temporary increase in EDA after the prompt), but the required support is completely different. If the learner experiences positive emotions, no further support is needed, but if the learner experiences negative emotions, support is required to lead the learner onto the growth pathway. Therefore, a measure is necessary to determine emotional valence and, thus, provide appropriate support. Nevertheless, my findings showed that if EDA constantly increases, the task is assessed negatively, and it is advisable to assist the learner. However, it must be considered that EDA can range from 2 to 20 microSiemens between individuals (Dawson et al., 2016). Therefore, defining a fixed value that indicates the well-being pathway is not possible. Consequently, only the change of EDA within an individual can be used (as presented in Study 1). Hence, further research is needed to investigate these changes in EDA to characterize the physiological appearance of both pathways more precisely.

In summary, the well-being pathway encompasses an exit strategy “identify increased EDA and detect the need for support” that leads to the growth pathway and, thus, high learning performance (see Figure 5, red path). In addition, an instruction on how the support can be generated is included. Here, SRL skills and learning preferences must be considered. Moreover, changes in EDA must be linked to navigation behavior to identify the learning process correctly. As a result, my findings can further improve computer-based learning and help educators and learners to understand learning processes based on evaluating process data.

7.2.2 The Technological Pedagogical Evaluation and Content Knowledge Model (TPEACK)

Next, I will refer to the TPACK model described in chapter 2.1. The TPACK model states that educators need knowledge across all three core competencies (TK, PK, CK; see Figure 1) and their intersections (TCK, PCK, TPK; Figure 1) to use digital media successfully.

However, no competency that includes knowledge about evaluating learning processes is described, although this is considered crucial to improve learning (e.g., Ahlan et al., 2014; Azevedo & Gašević, 2019; Eickelmann & Gerick, 2020; Hattie, 2017). Therefore, I want to close this research gap based on my findings. I expanded the TPACK model with the competency of evaluating learning processes, which I called "evaluation knowledge" (EK) as a part of the technological knowledge. Moreover, in this area, knowledge about evaluation methods (e.g., log file or psychophysiological data analyses) and their impact on learning success is located (see Figure 6).

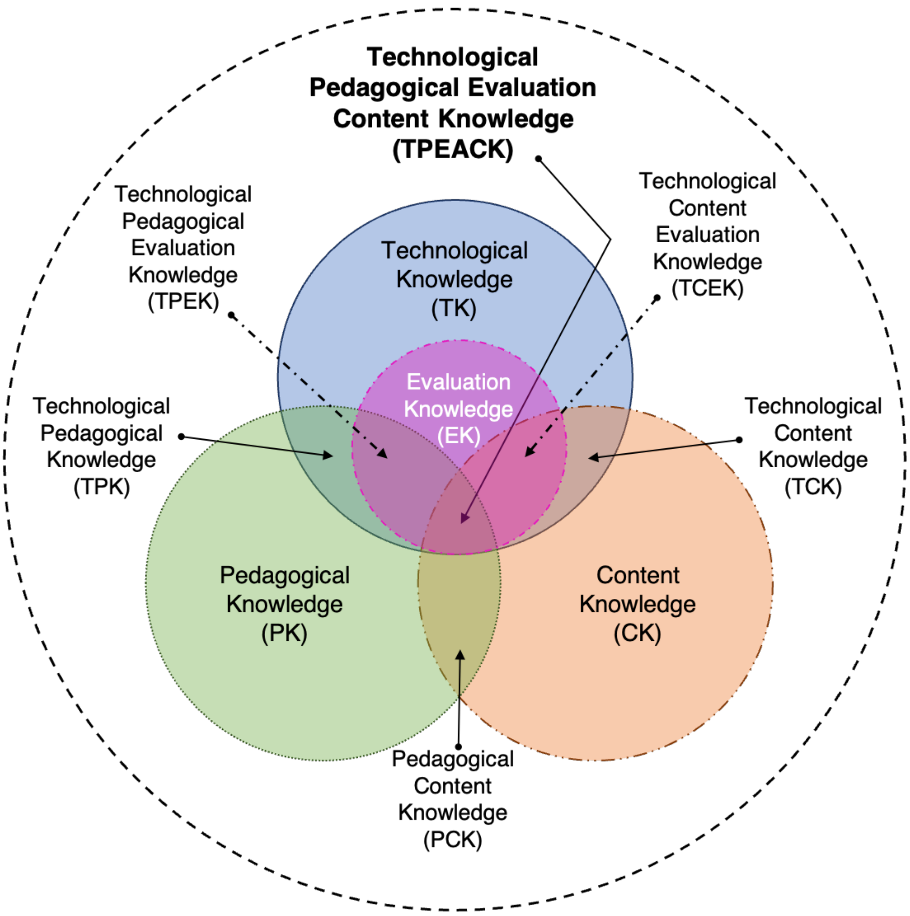


Figure 6. The Technical Pedagogical Evaluation and Content Knowledge Model (TPEACK), adapted from the TPACK model (Koehler & Mishra, 2009, p. 63).

The intersection with technological knowledge refers to the educators' understanding of analyzing, interpreting, and promoting the learning processes (ETK). The combination of evaluation and pedagogical knowledge addresses how to assess whether the CBLE fits the pedagogical methods. The intersection of the evaluation and content knowledge describes evaluating whether the CBLE adequately transmits the subject matter. In sum, I developed the TPEACK model, which demonstrates the importance of evaluation methods within the TK and thus examines whether the CBLE is being used successfully. Moreover, the TPEACK model illustrates that, in addition to the pedagogical, content-related, and technological competencies, the evaluation of learning processes must also be part of the repertoire of educators to implement CBLEs meaningfully. From there, considerable practical implications can be drawn, presented in the following chapter.

7.3 Practical implications

At this point, I want to apply my findings in practice. Not all educators know how to evaluate learning processes in CBLEs, due to their media literacy level, especially using process data (Eickelmann et al., 2019; Eickelmann & Gerick, 2020; Schneider et al., 2021). Therefore, workshops and training are necessary. To realize these workshops, the parent institution must provide the resources and digital infrastructure (e.g., allowing time for training, integration in the curriculum). Unfortunately, this is not yet the case in German schools. The ICILS study from 2018 (Eickelmann et al., 2019) revealed that only 25.9% of educators in Germany learned how to use digital media in their teacher training. The fact that this value is significantly below the international average (41.6%) is critical. Furthermore, less than 32% of educators participated in webinars, training, or courses on implementing and using digital media. However, 60.2% of educators in Germany state that they use digital media in their teaching at least once a week. Almost a quarter (23.2%) of educators in Germany use digital media daily. This discrepancy demonstrates that educators must be trained to use digital media successfully. Furthermore, using digital media to provide support rarely occurs in German classrooms. The reason could be that only 33.6% of educators are confident in using a learning management system (e.g., Moodle), and only 34.7% of educators are convinced that digital media can improve academic performance (Eickelmann et al., 2019). This shows that digital media has entered German educational institutions, but the knowledge about its use and benefits needs improvement.

Based on these facts, I want to introduce a top-down approach to improving the use of digital media in educational institutions. First, the school or the superordinate institution (e.g., ministry) must initiate appropriate changes at the school directly (e.g., providing tablets and functioning wireless networks) and in the curriculum (e.g., integration of mandatory workshops and training). In the next step, educators need further qualifications in their media literacy. Here, the main focus lies on motivation and acceptance of digital media. This consideration should lead educators to realize digital media's benefits and know how digital media (e.g., CBLEs) can be purposefully implemented in the individual teaching concept. Consequently, media literacy training and the implementation of digital media according to the curriculum must go hand in hand. In training media literacy, the dimensions of my TPEACK model (see Figure 6, chapter 7.2) must be considered and can serve as a framework. Here, all essential skills to use digital media successfully are presented (Koehler & Mishra, 2009).

In addition to educator training, learner acceptance of digital media must be addressed. Since most of today's students are digital natives and have grown up with digital media, it can be assumed that they are sufficiently media literate. However, Persada and colleagues (2019) found correlations between the dimensions of the *Unified Theory of Acceptance and Use of Technology* (Venkatesh et al., 2003), where three factors are described that impact the acceptance use of technologies. The highest correlation was found between the factor "facilitating conditions" and the behavioral intention to use digital media. This means that the acceptance and, thus, the intention to use digital media is exceptionally high if students are supported and have the required resources to use digital media (Persada et al., 2019; Venkatesh et al., 2003). This finding underpins that institutions and educators are responsible for successfully implementing digital media in teaching and learning. Therefore, my presented top-down approach is a fruitful concept to anchor digital media in educational institutions.

Furthermore, a third important stakeholder is affected by my findings. The results of the present research also influence developers of CBLEs. In Study 2, I demonstrated the relationship between learning success and navigation behavior. From this, a pattern of navigating through a CBLE to achieve high learning performance can be derived. Therefore, developers of CBLEs can use these insights to foster conducive navigation. It has been proven beneficial for learning to circle between texts and tasks (see chapter 6.2 and Manuscript B). Hence, linking text pages with corresponding tasks is a feasible example of promoting learning success. Another feature to facilitate the use of digital media (especially for complex CBLEs) can be walkthroughs. Here, the structure and usage of a CBLE can be explained, including exemplary application scenarios. Moreover, CBLE developers can offer workshops for learners and educators, which leads to optimal usage of the CBLE. A well-elaborated CBLE where these features are implemented is the TTE used in Study 2 (see chapter 5.2.3; Lewalter et al., 2018a). The TTE includes tutorials on user knowledge for educators and learners. Moreover, videos about the aim, content, and first steps are available, and tasks are linked within text pages.

7.3.1 The TPEACK model

Whereas Study 1 suggests more methodological and theoretical implications (see chapter 7.2), findings from Study 2 yield more practical outcomes. As presented in chapters 6.2, 7.1.2.1, and Manuscript B, log file data revealed learner groups that differ in their learning performance based on their navigation behavior. Hence, educators can identify different learner groups by evaluating log file data. This skill can foster competencies presented in the TPACK model (see chapter 2.1 and Figure 1). The intersection of pedagogical and content knowledge describes how to prepare specific contents to address different learner groups (see Figure 1; Koehler & Mishra, 2009). As a result, by evaluating log files, educators can identify and promote different learner groups by tailoring content to meet learners' individual needs. This specific competency is the core of my TPEACK model, presented in chapter 7.2 (see Figure 6). Here, all competencies of the model are combined. The educator knows how to use a CBLE (TK) to teach the content (CK) in a pedagogically appropriate way (PK) and knows how to evaluate the learning process (EK).

Nevertheless, the question of technical implementation arises. The focus of this research is real-time evaluation and support during computer-based learning. Since I explicitly refer to CBLEs, where learning is possible at any time, the educator must constantly be available, which is almost impossible. My solution for this issue is to develop an algorithm that automatically evaluates log files, identifies navigation behaviors, and provides individualized support using predefined prompts. The development of this kind of algorithm is technically feasible, as shown in Study 2 and prior research (Lim et al., 2023). Issuing individualized support seems more complicated at first. However, the educator can implement prepared prompts into the algorithm, which are sent according to the individual interaction of the learner with the CBLE (a detailed discussion can be seen in chapters 7.1.2.1 and 7.2). Here, the interaction and time spent on learning-relevant pages should be focused on (see Study 2; Hörmann & Bannert, 2016; Jeske et al., 2014; Narciss et al., 2007). For example, if the learner does not interact with the CBLE after a specific time, a prompt can be presented (e.g., after 15 minutes, send prompt 3) to encourage the learner. At best, the prompts are customized for each learner based on prior interactions and experiences. A practical example is the language learning app "Duolingo" (www.duolingo.com). Here, lessons that repeat the respective topics are prepared based on the learners' previous mistakes. This concept, as shown in Study 2 and Lim and colleagues (2023), demonstrates that it is technically realizable to adapt individual support to the learner in automatically real time.

While developing and implementing an algorithm in CBLEs to evaluate log files automatically is feasible, generating an algorithm for evaluating psychophysiological data is much more complex. Although the algorithm's prototypical command "if value x, send prompt y" would remain identical, there are several factors to consider for psychophysiological data. As described in chapter 5.1, EDA and HR can be affected by body movement (Potter & Bolls, 2012). Therefore, body movements are seen as artifacts and need to be removed before analyzing the data (see chapter 5.1.6). Considering the case of learners studying at home, which is mostly the case using CBLEs, these artifacts can be difficult to control, leading to a possible bias in the data (Potter & Bolls, 2012). Another issue would be that the learners need to attach cables to themselves, which can be error-prone without knowledge about physiological data collection. At this point, it becomes clear that implementing psychophysiological data collection in self-studying is very elaborate. Nevertheless, previous research included psychophysiological measures in learning settings using wearables (Malmberg, Haataja, et al., 2019; Malmberg, Järvelä, et al., 2019). These wearable electrodes are easy to apply and provide an practical approach to tracking psychophysiological measures (Goetz et al., 2022). However, these studies have been conducted in the classroom under supervision. Therefore, further research is needed on how psychophysiological measurements can be recorded and evaluated in self-study phases. My findings contributed to the understanding of psychophysiological responses to emotionally activating learning materials and, thus, achieving a more accurate evaluation of EDA and HR. However, more research is required on the technical implementation in self-study phases.

7.4 Limitations and future research

The limitations of my studies point the direction for future research. The first limitation addresses the results from Study 1. I demonstrated that the HR could not reveal the valence of the emotion due to the highly emotional activating learning material (for a detailed discussion, see chapter 7.1.1.1). Therefore, in future research, attention should be paid to the emotionality of the learning materials and, thus, whether HR is an appropriate measure for valence. In chapter 7.2, I state that if the learning materials contain highly emotional content, it is questionable whether HR can display the valence of emotions. However, especially when analyzing learning processes, the valence of emotions is a crucial factor for predicting learning success (Goetz et al., 2022; Loderer et al., 2020). Positive and negative emotions can increase EDA, but the type of support changes fundamentally. If the learner experiences excitement because he or she solved a task correctly, EDA increases (Boucsein, 2012; Malmberg et al., 2015). Since positive emotions benefit learning, no intervention is needed; it could even hinder learning (C. Huang, 2011). However, if the learner is frustrated due to failure, EDA also increases, and support is necessary (Pekrun & Stephens, 2012). In conclusion, identifying emotional valence is fundamental in providing adequate support and promoting learning (Malmberg et al., 2015).

As discussed before, the learning materials from Study 1 evoke negative activating emotions, although learners showed a significant increase in learning performance (cf. Boekaerts, 2011). Noteworthy is that the prior knowledge was relatively low due to the unfamiliar topic of the learning materials. Therefore, it is reasonable that a higher posttest score was achieved. However, future studies should prefer a more popular topic. Nevertheless, Goetz and colleagues (2022) stated that despite the evidence that positive emotions are beneficial and negative emotions are detrimental to learning, these connections might be more differentiated. Thus, more research is needed to clarify these nuances (Goetz et al., 2022).

Another limitation of Study 1 is the technical implementation and the practical integration of evaluating psychophysiological measurements in self-studying. As in recent years, psychophysiological measurements have been increasingly used in education; their popularity is conspicuous (e.g., Järvelä et al., 2021, 2023; Malmberg, Haataja, et al., 2019; Malmberg, Järvelä, et al., 2019). However, most studies took place inside the classroom under supervision. Since learning with CBLEs also gained importance, which mostly takes place outside the classroom, the use of psychophysiological measures in self-studying should be focused on in future research (Eickelmann & Gerick, 2020; Lajoie et al., 2019).

Regarding Study 2, the general limitation of log file data can be mentioned. In prior research, log files are labeled as “ontological flat” (Järvelä et al., 2021; Reimann et al., 2014) compared to fine-grained data as think-aloud protocols (Fan et al., 2022; Järvelä et al., 2021). However, log file data have been successfully used in research to evaluate several learning-related activities (e.g., SRL, individual learning strategies) and are especially fruitful regarding the flow and unfolding of learners’ activities (for a meta-analysis, see Guo, 2022). Moreover, log file data can provide objective insights into learners’ behavior and psychological processes associated with, for example, changes in attention or effort that are difficult to detect otherwise (Winne, 2010).

In addition, log files and psychophysiological data are multifaceted and rich data (Järvelä et al., 2021). Therefore, it is possible to extract diverse information. For example, EDA data provide tonic (large-scale) and phasic (small-scale) responses (Dawson et al., 2016). Depending on the research aim, it is helpful to consider the overall learning session or a single reaction to a stimulus. However, large-scale information can be unspecific and inconclusive. For example, the learner’s overall duration in the CBLE does not provide detailed information about the navigation behavior. Also, analyzing an overall mean of EDA undermines the benefit of psychophysiological data for visualizing unfolding processes. Therefore, it is valuable to analyze specific areas, for example, a physiological reaction to a prompt (see chapter 7.2.1). Moreover, using physiological data to identify trigger events for learning activities makes it easier to interpret the interplay between physiological responses and learning processes (e.g., Järvelä et al., 2023). Regarding log files, it is helpful to label and categorize pages to receive meaningful information (e.g., learning-relevant, orienting, learning-irrelevant; see Study 2). Therefore, future studies must consider the granularity of the desired information in order to obtain significant results.

A profitable opportunity for future research would be to examine my results in the long-run. An exploratory approach could be to investigate whether psychophysiological reactions to individualized prompts or features of a CBLE habituate over time (Boucsein, 2012). Hence, future research could implement my findings from the laboratory into the field and examine the long-term effects of psychophysiological responses to individualized support and its relation to navigation behavior as well as learning performance.

7.5 Conclusion

In conclusion, the present thesis aims to promote computer-based learning by investigating learning processes through process data. By identifying detrimental and beneficial patterns in EDA and HR and log file data, support can be provided in real time. Moreover, monitoring and evaluating individual learning activities and support (e.g., prompts) can be tailored to the learner and task. Based thereon, computer-based learning can be improved without disrupting the learning process. Therefore, this work seeks to determine how psychophysiological data and navigation behavior can sufficiently indicate learning processes and predict learning performance. In order to investigate the manifestation of learning processes in process data, two approaches were examined.

The first approach researched the physiological appearance of academic emotions and their impact on learning outcomes. Results demonstrated that negative activating academic emotions were expressed in high EDA and HR. The most meaningful finding was that EDA significantly predicted learning outcomes. Moreover, depending on learning performance (high, middle, low), the psychophysiological behavior differed, meaning high learning outcomes were associated with low EDA and HR. Interestingly, boredom decreased after learning and was expressed through low HR. However, HR was no significant predictor for learning performance. Therefore, more research is needed to clarify the relation between HR as a valence measure and highly emotionally activating learning content.

The second approach aimed to investigate the extent to which log files can identify learning processes in CBLEs. Moreover, the impact of navigation behavior on learning performance was examined. Therefore, a dendrogram was generated, which revealed two learner groups that differed significantly in their navigation behavior and learning outcomes. High performers were characterized by active interaction, significantly higher duration of stay, and learning outcomes. Furthermore, a process model demonstrated that the high performers showed a meaningful loop pattern of navigating between text pages and tasks.

In conclusion, educators can indicate successful learning by focusing on the duration on learning-relevant pages, active interaction, and low EDA levels. Based on these measures, individualized support can be provided. Moreover, emotional responses and navigation behavior can be used to derive design principles to shape the learning experience. In sum, linking log files and physiological data allow fine-grained insights into learning processes. Thus, my implications provide a valuable framework for interpreting process data. Future research is needed to apply the theoretical models I derived and to investigate physiological responses to the CBLE over the long term.

8 References

Adewoye, S. E., & Olaseni, A. O. (2022). A Systematic Review of Factors That Influences the Efficacy of Computer Based Learning. *Journal of Education, Teaching, and Learning*, *7*(2), 166–175.

Ahlan, A., Atanda, B. N., & Isah, Y. (2014). Information Technology Enhances Students' Academic Performance: A Case Of University Of Ilorin. *The Online Journal of Distance Education and E-Learning*, *2*(2), 6.

Arguel, A., Lockyer, L., Lipp, O. V., Lodge, J. M., & Kennedy, G. (2017). Inside Out: Detecting Learners' Confusion to Improve Interactive Digital Learning Environments. *Journal of Educational Computing Research*, *55*(4), 526–551. https://doi.org/10.1177/0735633116674732

Arnold, P., Kilian, L., Thillosen, A. M., & Zimmer, G. (Eds.). (2015). *Handbuch E-Learning: Lehren und Lernen mit digitalen Medien* (4., erw. Aufl). Bertelsmann.

Azevedo, R., & Gašević, D. (2019). Analyzing Multimodal Multichannel Data about Self-Regulated Learning with Advanced Learning Technologies: Issues and Challenges. *Computers in Human Behavior*, *96*, 207–210. https://doi.org/10.1016/j.chb.2019.03.025

Azevedo, R., Harley, J., Trevors, G., Duffy, M., Feyzi-Behnagh, R., Bouchet, F., & Landis, R. (2013). Using Trace Data to Examine the Complex Roles of Cognitive, Metacognitive, and Emotional Self-Regulatory Processes During Learning with Multi-agent Systems. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (Vol. 28, pp. 427–449). Springer New York. https://doi.org/10.1007/978-1-4419-5546-3_28

Azevedo, R., Taub, M., & Mudrick, N. V. (2017). Understanding and Reasoning about Real-Time Cognitive, Affective, and Metacognitive Processes to Foster Self-Regulation with Advanced Learning Technologies. In D. H. Schunk & J. A. Greene (Eds.), *Handbook of Self-Regulation of Learning and Performance* (2nd ed., pp. 254–270). Routledge. https://doi.org/10.4324/9781315697048-17

Baker, R. S. J. d., D'Mello, S. K., Rodrigo, Ma. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223–241. https://doi.org/10.1016/j.ijhcs.2009.12.003

Bannert, M. (2003). Effekte metakognitiver Lernhilfen auf den Wissenserwerb in vernetzten Lernumgebungen. *Zeitschrift für Pädagogische Psychologie, 17*(1), 13–25. https://doi.org/10.1024//1010-0652.17.1.13

Bannert, M. (2006). Effects of Reflection Prompts When Learning with Hypermedia. *Journal of Educational Computing Research, 35*(4), 359–375. https://doi.org/10.2190/94V6-R58H-3367-G388

Bannert, M. (2009). Promoting Self-Regulated Learning Through Prompts. *Zeitschrift Für Pädagogische Psychologie*, *23*(2), 139–145. https://doi.org/10.1024/1010-0652.23.2.139

Bannert, M., & Reimann, P. (2012). Supporting self-regulated hypermedia learning through prompts. *Instructional Science*, *40*(1), 193–211. https://doi.org/10.1007/s11251-011-9167-4

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, *9*(2), 161–185. https://doi.org/10.1007/s11409-013-9107-6

Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior*, *52*, 293–306. https://doi.org/10.1016/j.chb.2015.05.038

Berntson, G. G., Quigley, K. S., Norman, G. J., & Lozano, D. L. (2017). Cardiovascular Psychophysiology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 183–216). Cambridge University Press. https://doi.org/10.1017/9781107415782.009

BIOPAC Systems Inc. (2019). *Biopac Student Lab* (4.1) [Computer software].

Bloom, B.S. (Ed.), Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain.* New York: David McKay.

Boekaerts, M. (1999). Motivated learning: Studying student * situation transactional units. *European Journal of Psychology of Education*, *14*(1), 41-55.

https://doi.org/10.1007/BF03173110

Boekaerts, M. (2011). Emotions, emotion regulation, and self-regulation of learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 408–425). Routledge/Taylor & Francis Group.

Bolls, P. D., Lang, A., & Potter, R. F. (2001). The Effects of Message Valence and Listener Arousal on Attention, Memory, and Facial Muscular Responses to Radio Advertisements. *Communication Research*, *28*(5), 627–651. https://doi.org/10.1177/009365001028005003

Boucsein, W. (2012). *Electrodermal activity* (2nd ed.). Springer.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., roth, W. T., Dawson, M. E., Filion, D. L., & Society for Psychophysiological Research Ad Hoc Committee on Electrodermal, M. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Bruhn, D., & Wollenteit, U. (2018). Konventionelle Schweinehaltung und Tierschutzgesetz. *Natur und Recht*, *40*(3), 160–169. https://doi.org/10.1007/s10357-018-3310-9

Butcher, K. R. (2014). The Multimedia Principle. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 174–205). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.010

Cacioppo, J. T., & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *American Psychologist*, *45*(1), 16–28. https://doi.org/10.1037/0003-066X.45.1.16

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2016). Strong Inference in Psychophysiological Science. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 3–15). Cambridge University Press. https://doi.org/10.1017/9781107415782.001

Cerezo, R., Bogarín, A., Esteban, M., & Romero, C. (2020). Process mining for self-regulated learning assessment in e-learning. *Journal of Computing in Higher Education*, *32*(1), 74–88. https://doi.org/10.1007/s12528-019-09225-y

Cheok, M. L., Wong, S. L., & Ayub, A. F. (2017). Teachers' Perceptions of E-Learning in Malaysian Secondary Schools. *Malaysian Online Journal of Educational Technology, 5*(2), 14.

Churches, A. (2008). Bloom's taxonomy blooms digitally. *Tech & Learning*, *1*, 1–6.

Cranford, K. N., Tiettmeyer, J. M., Chuprinko, B. C., Jordan, S., & Grove, N. P. (2014). Measuring Load on Working Memory: The Use of Heart Rate as a Means of Measuring Chemistry Students' Cognitive Load. *Journal of Chemical Education*, *91*(5), 641–647. https://doi.org/10.1021/ed400576n

Dawson, M. E., Schell, A. M., & Filion, D. L. (2016). The Electrodermal System. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 217–243). Cambridge University Press. https://doi.org/10.1017/9781107415782.010

Dindar, M., Malmberg, J., Järvelä, S., Haataja, E., & Kirschner, P. A. (2019). Matching self-reports with electrodermal activity data: Investigating temporal changes in self-regulated learning. *Education and Information Technologies*, *25*(3), 1785–1802. https://doi.org/10.1007/s10639-019-10059-5

D'Mello, S., & Graesser, A. (2014). Confusion and its dynamics during device comprehension with breakdown scenarios. *Acta Psychologica*, *151*, 106–116. https://doi.org/10.1016/j.actpsy.2014.06.005

Dolak, R. (2019). Using Process Mining Techniques to Discover Student's Activities, Navigation Paths, and Behavior in LMS Moodle. In L. Rønningsbakk, T.-T. Wu, F. E. Sandnes, & Y.-M. Huang (Eds.), *Innovative Technologies and Learning* (Vol. 11937, pp. 129–138). Springer International Publishing. https://doi.org/10.1007/978-3-030-35343-8_14

Donker, M. H., van Gog, T., & Mainhard, M. T. (2018). A Quantitative Exploration of Two Teachers with Contrasting Emotions: Intra-Individual Process Analyses of Physiology and Interpersonal Behavior. *Frontline Learning Research*, 162–185. https://doi.org/10.14786/flr.v6i3.372

Duffy, M. C., & Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior*, *52*, 338–348.

Duffy, M. C., Lajoie, S. P., Pekrun, R., & Lachapelle, K. (2020). Emotions in medical education: Examining the validity of the Medical Emotion Scale (MES) across authentic medical learning environments. *Learning and Instruction*, *70*, 101150. https://doi.org/10.1016/j.learninstruc.2018.07.001

Eickelmann, B., Bos, W., Gerick, J., Goldhammer, F., Schaumburg, H., Schwippert, K., Senkbeil, M., & Vahrenhold, J. (Eds.). (2019). *ICILS 2018 #Deutschland: Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking*. Waxmann.

Eickelmann, B., & Gerick, J. (2020). Lernen mit digitalen Medien. Zielsetzungen in Zeiten von Corona und unter besonderer Berücksichtigung von sozialen Ungleichheiten. In D. Fickermann & B. Edelstein (Eds.), *„Langsam vermisse ich die Schule ...“* (pp. 153–162). Waxmann Verlag GmbH. https://doi.org/10.31244/9783830992318.09

Engelmann, K., & Bannert, M. (2021). Analyzing temporal data for understanding the learning process induced by metacognitive prompts. *Temporal and Adaptive Process of Regulated Learning -What Can Multimodal Data Tell?*, *72*, 101205. https://doi.org/10.1016/j.learninstruc.2019.05.002

Eteläpelto, A., Kykyri, V.-L., Penttonen, M., Hökkä, P., Paloniemi, S., Vähäsantanen, K., Eteläpelto, T., & Lappalainen, V. (2018). A multi-componential methodology for exploring emotions in learning. *Frontline Learning Research*, 6–36. https://doi.org/10.14786/flr.v6i3.379

Fan, Y., Lim, L., van der Graaf, J., Kilgour, J., Raković, M., Moore, J., Molenaar, I., Bannert, M., & Gašević, D. (2022). Improving the measurement of self-regulated learning using multi-channel data. *Metacognition and Learning*. https://doi.org/10.1007/s11409-022-09304-z

Fluxicon BV. (2022). *Disco* (3.0.0) [Computer software].

Fortin-Cote, A., Beaudin-Gagnon, N., Campeau-Lecours, A., Tremblay, S., & Jackson, P. L. (2019). Affective Computing Out-of-The-Lab: The Cost of Low Cost. *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, 4137–4142. https://doi.org/10.1109/SMC.2019.8914646

Goedhart, N. S., Blignaut-van Westrhenen, N., Moser, C., & Zweekhorst, M. B. M.

(2019). The flipped classroom: Supporting a diverse group of students in their learning. *Learning Environments Research*, *22*(2), 297–310. https://doi.org/10.1007/s10984-019-09281-2

Goetz, T., Frenzel, A. C., Stockinger, K., Lipnevich, A. A., Stempfer, L., & Pekrun, R. (2022). *Emotions in Education*. *4*, 36.

Goetz, T., & Hall, N. C. (2013). Academic Boredom. In *International Handbook of Emotions in Education*. Routledge. https://doi.org/10.4324/9780203148211.ch16

Goldman, S. R. (2009). Explorations of relationships among learners, tasks, and learning. *Learning and Instruction*, *19*(5), 451–454. https://doi.org/10.1016/j.learninstruc.2009.02.006

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125. https://doi.org/10.1007/s40881-015-0004-4

Grewenig, E., Lergetporer, P., Werner, K., Woessmann, L., & Zierow, L. (2021). COVID-19 and educational inequality: How school closures affect low-and high-achieving students. *European Economic Review*, *140*, 103920.

Guo, L. (2022). Using metacognitive prompts to enhance self-regulated learning and learning outcomes: A meta-analysis of experimental studies in computer-based learning environments. *Journal of Computer Assisted Learning*, *38*(3), 811–832. https://doi.org/10.1111/jcal.12650

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-physiological measures for assessing cognitive load. *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, 301–310. https://doi.org/10.1145/1864349.1864395

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, *2*(2–3), 107–124. https://doi.org/10.1007/s11409-007-9016-7

Hattie, J. (2017). Visible learning plus: 250+ influences on student achievement. *Visible learning plus.*

Hong, J.-C., Liu, Y., Liu, Y., & Zhao, L. (2021). High School Students' Online Learning Ineffectiveness in Experimental Courses During the COVID-19 Pandemic.

Frontiers in Psychology, *12*, 3766. https://doi.org/10.3389/fpsyg.2021.738695

Hörmann, M., & Bannert, M. (2016). *Visualising and analysing self-regulated learning by recording the learning process: Introducing peripheral data as a better alternative to screen recordings and log-files*. EARLI SIG 16 Metacognition.

Hoss, T., Ancina, A., & Kaspar, K. (2021). Forced Remote Learning During the COVID-19 Pandemic in Germany: A Mixed-Methods Study on Students' Positive and Negative Expectations. *Frontiers in Psychology*, *12*, 3469. https://doi.org/10.3389/fpsyg.2021.642616

Huang, C. (2011). Achievement Goals and Achievement Emotions: A Meta-analysis. *Educational Psychology Review*, *23*(3), 359–388. https://doi.org/10.1007/s10648-011-9155-x

Huang, L., & Lajoie, S. P. (2021). Process analysis of teachers' self-regulated learning patterns in technological pedagogical content knowledge development. *Computers & Education*, *166*, 104169. https://doi.org/10.1016/j.compedu.2021.104169

Huber, K., & Bannert, M. (2022). What happens to your body during learning with computer-based environments? Exploring negative academic emotions using psychophysiological measurements. *Journal of Computers in Education*. https://doi.org/10.1007/s40692-022-00228-w

Huber, S. G., & Helm, C. (2020). COVID-19 and schooling: Evaluation, assessment and accountability in times of crises—reacting quickly to explore key issues for policy, practice and research with the school barometer. *Educational Assessment, Evaluation and Accountability*, *32*(2), 237–270. https://doi.org/10.1007/s11092-020-09322-y

IBM Corp. (2020). *IBM SPSS Statistics for Windows* (26) [Computer software].

Ijsselsteijn, de Ridder, H., Freeman, J., & Avons, S. E. (2000). Presence: Concept, determinants and measurement. *Human Vision and Electronic Imaging*, *3959*, 520–529.

iMotions. (2019). *iMotions Research Software* (8.1) [Computer software]. https://imotions.com

Izard, C. E. (2010). The Many Meanings/Aspects of Emotion: Definitions, Functions, Activation, and Regulation. *Emotion Review*, *2*(4), 363–370.

https://doi.org/10.1177/1754073910374661

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2019). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*, 101203. https://doi.org/10.1016/j.learninstruc.2019.04.004

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2021). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*, *72*, 101203. https://doi.org/10.1016/j.learninstruc.2019.04.004

Järvelä, S., Nguyen, A., Vuorenmaa, E., Malmberg, J., & Järvenoja, H. (2023). Predicting regulatory activities for socially shared regulation to optimize collaborative learning. *Computers in Human Behavior*, *144*, 107737. https://doi.org/10.1016/j.chb.2023.107737

JASP Team. (2020). *JASP* (0.12.2) [Computer software].

Jennings, J. R., & Allen, B. (2016). Methodology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 583–611). Cambridge University Press. https://doi.org/10.1017/9781107415782.027

Jeske, D., Backhaus, J., & Stamov Roßnagel, C. (2014). Self-regulation during e-learning: Using behavioural evidence from navigation log files: Self-regulation during e-learning. *Journal of Computer Assisted Learning*, *30*(3), 272–284. https://doi.org/10.1111/jcal.12045

Jonkisz, E., Moosbrugger, H., & Brandt, H. (2012). Planung und Entwicklung von Tests und Fragebogen. In *Testtheorie und Fragebogenkonstruktion* (pp. 27–74). Springer.

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G. F., McClure, S. M., Wang, J. T., & Camerer, C. F. (2008). The Wick in the Candle of Learning: Epistemic Curiosity Activates Reward Circuitry and Enhances Memory. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1308286

Koehler, M., & Mishra, P. (2009). What is Technological Pedagogical Content Knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, *9*(1), 60–70.

Kovanović, V., Joksimović, S., Poquet, O., Hennis, T., De Vries, P., Hatala, M., Dawson, S., Siemens, G., & Gašević, D. (2019). Examining communities of inquiry in Massive Open Online Courses: The role of study strategies. *The Internet and Higher Education*, *40*, 20–43. https://doi.org/10.1016/j.iheduc.2018.09.001

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der" Positive and Negative Affect Schedule"(PANAS). *Diagnostica-Gottingen-*, *42*, 139–156.

Laarni, J., Ravaja, N., Saari, T., Böcking, S., Hartmann, T., & Schramm, H. (2015). Ways to measure spatial presence: Review and future directions. In *Immersed in Media* (pp. 139–185). Springer.

Lajoie, S. P., Pekrun, R., Azevedo, R., & Leighton, J. P. (2019). Understanding and measuring emotions in technology-rich learning environments. *Learning and Instruction*, 101272. https://doi.org/10.1016/j.learninstruc.2019.101272

Lang, A. (2014). What can the heart tell us about thinking? In *Measuring psychological responses to media messages* (pp. 111–124). Routledge.

Lang, A., Potter, R., & Bolls, P. (2009). Where psychophysiology meets the media: Taking the effects out of mass media research. *Media Effects: Advances in Theory and Research*, 185–206.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, *1*(39–58), 3.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology*, *30*(3), 261–273. https://doi.org/10.1111/j.1469-8986.1993.tb03352.x

Leiner, D. J. (2019). *SoSci Survey* (3.1.06) [Computer software].

Leppert, K., Koch, B., Brähler, E., & Strauß, B. (2008). *Die Resilienzskala (RS) – Überprüfung der Langform RS-25 und einer Kurzform RS-13*. 18.

Levenson, R. W., Lwi, S. J., Brown, C. L., Ford, B. Q., Otero, M. C., & Verstaen, A. (2016). Emotion. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 444–464). Cambridge University Press. https://doi.org/10.1017/9781107415782.020

Lewalter, D., Schiffhauer, S., Richter-Gebert, J., Bannert, M., Engl, A.-T., Maahs, M., Reißner, M., Ungar, P., & von Wachter, J.-K. (2018a). *Toolbox Lehrerbildung* [Computer software]. https://toolbox.edu.tum.de/

Lewalter, D., Schiffhauer, S., Richter-Gebert, J., Bannert, M., Engl, A.-T., Maahs, M., Reißner, M., Ungar, P., & von Wachter, J.-K. (2018b). Toolbox Lehrerbildung: Berufsfeldbezogene Vernetzung von Fach, Fachdidaktik und Bildungswissenschaft. *Kohärenz in Der Universitären Lehrerbildung*, 331–353.

Lewalter, D., Schneeweiss, A., Richter-Gebert, J., Huber, K., & Bannert, M. (2022). Mit Unterrichtsvideos praxisnah und disziplinverbindend lehren und lernen. *Lehren Und Forschen Mit Videos in Der Lehrkräftebildung*, 125.

Lewalter, D., Titze, S., Bannert, M., & Richter-Gebert, J. (2020). Lehrer*innenbildung digital und disziplinverbindend. Die Toolbox Lehrerbildung. *Journal für LehrerInnenbildung*, 76–84. https://doi.org/10.35468/jlb-02-2020_06

Lim, L., Bannert, M., van der Graaf, J., Molenaar, I., Fan, Y., Kilgour, J., Moore, J., & Gašević, D. (2021). Temporal Assessment of Self-Regulated Learning by Mining Students' Think-Aloud Protocols. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/article/10.3389/fpsyg.2021.749749

Lim, L., Bannert, M., van der Graaf, J., Singh, S., Fan, Y., Surendrannair, S., Rakovic, M., Molenaar, I., Moore, J., & Gašević, D. (2023). Effects of real-time analytics-based personalized scaffolds on students' self-regulated learning. *Computers in Human Behavior*, *139*, 107547. https://doi.org/10.1016/j.chb.2022.107547

Linnenbrink-Garcia, L., & Pekrun, R. (2011). Students' emotions and academic engagement: Introduction to the special issue. *Contemporary Educational Psychology*, *36*(1), 1–3. https://doi.org/10.1016/j.cedpsych.2010.11.004

Loderer, K., Pekrun, R., & Lester, J. C. (2020). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction*, *70*, 101162. https://doi.org/10.1016/j.learninstruc.2018.08.002

Magno, C. (2011). Validating the Academic Self-regulated Learning Scale with the

Motivated Strategies for Learning Questionnaire (MSLQ) and Learning and Study Strategies Inventory (LASSI). *The International Journal of Educational and Psychological Assessment*, *7*.

Malmberg, J., Haataja, E., Seppänen, T., & Järvelä, S. (2019). Are we together or not? The temporal interplay of monitoring, physiological arousal and physiological synchrony during a collaborative exam. *International Journal of Computer-Supported Collaborative Learning*, *14*(4), 467–490. https://doi.org/10.1007/s11412-019-09311-4

Malmberg, J., Järvelä, S., Holappa, J., Haataja, E., Huang, X., & Siipo, A. (2019). Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Computers in Human Behavior*, *96*, 235–245. https://doi.org/10.1016/j.chb.2018.06.030

Malmberg, J., Järvelä, S., Järvenoja, H., & Panadero, E. (2015). Promoting socially shared regulation of learning in CSCL: Progress of socially shared regulation among high- and low-performing groups. *Computers in Human Behavior*, *52*, 562–572. https://doi.org/10.1016/j.chb.2015.03.082

Malmberg, J., Järvenoja, H., & Järvelä, S. (2010). Tracing elementary school students' study tactic use in gStudy by examining a strategic and self-regulated learning. *Computers in Human Behavior*, *26*(5), 1034–1042. https://doi.org/10.1016/j.chb.2010.03.004

Matcha, W., Gašević, D., Uzir, N. A., Jovanović, J., & Pardo, A. (2019). Analytics of Learning Strategies: Associations with Academic Performance and Feedback. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 461–470. https://doi.org/10.1145/3303772.3303787

Matomo.org. (2022). *Matomo* [Computer software].

Mayer, R. E. (2014). Cognitive Theory of Multimedia Learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

McLaughlin, J. E., Roth, M. T., Glatt, D. M., Gharkholonarehe, N., Davidson, C. A., Griffin, L. M., Esserman, D. A., & Mumper, R. J. (2014). The Flipped Classroom: A Course Redesign to Foster Learning and Engagement in a Health Professions School.

Academic Medicine, *89*(2), 236–243. https://doi.org/10.1097/ACM.0000000000000086

McLaughlin, T., & Yan, Z. (2017). Diverse delivery methods and strong psychological benefits: A review of online formative assessment. *Journal of Computer Assisted Learning*, *33*(6), 562–574. https://doi.org/10.1111/jcal.12200

Mingorance Estrada, Granda Vera, Rojas Ruiz, & Alemany Arrebola. (2019). Flipped Classroom to Improve University Student Centered Learning and Academic Performance. *Social Sciences*, *8*(11), 315. https://doi.org/10.3390/socsci8110315

Molenaar, I., Mooij, S. D., Azevedo, R., Bannert, M., Järvelä, S., & Gašević, D. (2023). Measuring self-regulated learning and the role of AI: Five years of research using multimodal multichannel data. *Computers in Human Behavior*, *139*, 107540. https://doi.org/10.1016/j.chb.2022.107540

Moors, A. (2009). A review of theories concerned with emotion causation. *Cognition and Emotion*, *23*, 625–662.

Moos, D. C., & Azevedo, R. (2009). Learning With Computer-Based Learning Environments: A Literature Review of Computer Self-Efficacy. *Review of Educational Research*, *79*(2), 576–600. https://doi.org/10.3102/0034654308326083

Narciss, S., Proske, A., & Koerndle, H. (2007). Promoting self-regulated learning in web-based learning environments. *Computers in Human Behavior*, *23*(3), 1126–1144. https://doi.org/10.1016/j.chb.2006.10.006

Paans, C., Molenaar, I., Segers, E., & Verhoeven, L. (2020). Children's Macro-Level Navigation Patterns in Hypermedia and their Relation with Task Structure and Learning Outcomes. *Frontline Learning Research*, *8*(1), Article 1. https://doi.org/10.14786/flr.v8i1.473

Palomba, D., Angrilli, A., & Mini, A. (1997). Visual evoked potentials, heart rate responses and memory to emotional pictorial stimuli. *International Journal of Psychophysiology*, *27*(1), 55–67. https://doi.org/10.1016/S0167-8760(97)00751-4

Patel, D. B. (2013). Construction and effectiveness of computer aided instruction (CAI) programme for the units of science and technology of standard VIII. *International Journal of Research in Education*, *2*(1), 112–115.

Payne, A. F. H., Schell, A. M., & Dawson, M. E. (2016). Lapses in skin conductance responding across anatomical sites: Comparison of fingers, feet, forehead, and wrist: Lapses in skin conductance. *Psychophysiology*, *53*(7), 1084–1092. https://doi.org/10.1111/psyp.12643

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R. (2020). Self-Report is Indispensable to Assess Students' Learning. *Frontline Learning Research*, *8*(3), Article 3. https://doi.org/10.14786/flr.v8i3.637

Pekrun, R., Cusack, A., Murayama, K., Elliot, A. J., & Thomas, K. (2014). The power of anticipated feedback: Effects on students' achievement goals and achievement emotions. *Learning and Instruction*, *29*, 115–124. https://doi.org/10.1016/j.learninstruc.2013.09.002

Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: Exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, *102*(3), 531–549. https://doi.org/10.1037/a0019243

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist*, *37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors.* (pp. 3–31). American Psychological Association. https://doi.org/10.1037/13274-001

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: The Epistemically-Related Emotion Scales. *Cognition and*

Emotion, *31*(6), 1268–1276. https://doi.org/10.1080/02699931.2016.1204989

Persada, S. F., Miraja, B. A., & Nadlifatin, R. (2019). Understanding the Generation Z Behavior on D-Learning: A Unified Theory of Acceptance and Use of Technology (UTAUT) Approach. *International Journal of Emerging Technologies in Learning (iJET)*, *14*(05), 20. https://doi.org/10.3991/ijet.v14i05.9993

Pieger, E., & Bannert, M. (2018). Differential effects of students' self-directed metacognitive prompts. *Computers in Human Behavior*, *86*, 165–173. https://doi.org/10.1016/j.chb.2018.04.022

Pinel, J. P. J., & Pauli, P. (2012). *Biopsychologie* (8., aktualisierte Auflage). Pearson, Higher Education.

Potter, R. F., & Bolls, P. (2012). *Psychophysiological measurement and meaning: Cognitive and emotional processing of media*. Routledge.

Reimann, P., Markauskaite, L., & Bannert, M. (2014). e-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology*, *45*(3), 528–540. https://doi.org/10.1111/bjet.12146

Reisenzein, R. (2007). What is a definition of emotion? And are emotions mental-behavioral processes? *Social Science Information*, *46*(3), 424–428. https://doi.org/10.1177/05390184070460030110

Scherer, K. R., Schorr, A., & Johnstone, T. (2001). *Appraisal Processes in Emotion: Theory, Methods, Research*. Oxford University Press.

Schmidt-Atzert, L. (2009). Kategoriale und dimensionale Modelle. *Handbuch Der Allgemeinen Psychologie–Motivation Und Emotion*, 571–576.

Schneider, R., Sachse, K. A., Schipolowski, S., & Enke, F. (2021). Teaching in Times of COVID-19: The Evaluation of Distance Teaching in Elementary and Secondary Schools in Germany. *Frontiers in Education*, *6*, 702406. https://doi.org/10.3389/feduc.2021.702406

Schoor, C., & Bannert, M. (2012). Exploring regulatory processes during a computer-supported collaborative learning task using process mining. *Computers in Human Behavior*, *28*(4), 1321–1331. https://doi.org/10.1016/j.chb.2012.02.016

Schunk, D. H., & Greene, J. A. (Eds.). (2018). *Handbook of Self-Regulation of*

Learning and Performance (2nd eds.). Routledge, Taylor & Francis Group.

Sedlmeier, P., & Renkewitz, F. (2018). *Forschungsmethoden und Statistik für Psychologen und Sozialwissenschaftler* (3., aktualisierte und erweiterte Auflage). Pearson.

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*. https://doi.org/10.3389/fpubh.2017.00258

Siadaty, M., Gašević, D., & Hatala, M. (2016). Measuring the impact of technological scaffolding interventions on micro-level processes of self-regulated workplace learning. *Computers in Human Behavior*, *59*, 469–482. https://doi.org/10.1016/j.chb.2016.02.025

Sjak-Shie, E. E. (2019). *PhysioData Toolbox* (0.5) [Computer software]. https://PhysioDataToolbox.leidenuniv.nl

Slater, M. (2002). Presence and The Sixth Sense. *Presence: Teleoperators and Virtual Environments*, *11*(4), 435–439. https://doi.org/10.1162/105474602760204327

Sonnenberg, C., & Bannert, M. (2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1). https://doi.org/10.18608/jla.2015.21.5

Sonnenberg, C., & Bannert, M. (2016). Evaluating the Impact of Instructional Support Using Data Mining and Process Mining: A Micro-Level Analysis of the Effectiveness of Metacognitive Prompts. *Journal of Educational Data Mining*, *8*(2), 51–83.

Sonnenberg, C., & Bannert, M. (2019). Using Process Mining to examine the sustainability of instructional support: How stable are the effects of metacognitive prompting on self-regulatory behavior? *Computers in Human Behavior*, *96*, 259–272. https://doi.org/10.1016/j.chb.2018.06.003

Statistisches Bundesamt. (2020). *Digitales Lernen nimmt stark zu*. statista. https://de.statista.com/infografik/23712/nutzung-digitaler-lernangebote-in-deutschland/

Sypsas, A., & Kalles, D. (2022). Reviewing Process Mining Applications and Techniques in Education. *International Journal of Artificial Intelligence &*

Applications, *13*(1), 83–102. https://doi.org/10.5121/ijaia.2022.13106

Thompson, K., & Markauskaite, L. (2014). Identifying Group Processes and Affect in Learners: A Holistic Approach to Assessment in Virtual Worlds in Higher Education. In S. Kennedy-Clark, K. Everett, P. Wheeler, & R. Tennyson (Eds.), *Cases on the Assessment of Scenario and Game-Based Virtual Worlds in Higher Education*. IGI Global.

Titz, W. (2001). *Emotionen von Studierenden in Lernsituationen: Explorative Analysen und Entwicklung von Selbstberichtskalen*. Waxmann.

Titze, S., Schneeweiss, A., & Lewalter, D. (2021). Die Toolbox Lehrerbildung: Eine Lernplattform für die Professionalisierung von Lehramtsstudierenden im digitalen Zeitalter. *E-Teaching.Org-Artikel*.

Tivian XI GmbH. (2022). *Unipark.com* [Computer software].

Tugade, M. M., & Fredrickson, B. L. (2004). Resilient Individuals Use Positive Emotions to Bounce Back From Negative Emotional Experiences. *Journal of Personality and Social Psychology*, *86*(2), 320–333. https://doi.org/10.1037/0022-3514.86.2.320

van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review*, *28*, 100281. https://doi.org/10.1016/j.edurev.2019.05.003

van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2020). Self-regulated learning support in flipped learning videos enhances learning outcomes. *Computers & Education*, *158*, 104000. https://doi.org/10.1016/j.compedu.2020.104000

van Alten, D. C. D., Phielix, C., Janssen, J., & Kester, L. (2021). Secondary students' online self-regulated learning during flipped learning: A latent profile analysis. *Computers in Human Behavior*, *118*, 106676. https://doi.org/10.1016/j.chb.2020.106676

van der Graaf, J., Lim, L., Fan, Y., Kilgour, J., Moore, J., Gašević, D., Bannert, M., & Molenaar, I. (2022). The Dynamics Between Self-Regulated Learning and Learning Outcomes: An Exploratory Approach and Implications. *Metacognition and Learning*. https://doi.org/10.1007/s11409-022-09308-9

Van der Kleij, F. M., Feskens, R. C. W., & Eggen, T. J. H. M. (2015). Effects of

Feedback in a Computer-Based Learning Environment on Students' Learning Outcomes: A Meta-Analysis. *Review of Educational Research*, *85*(4), 475–511. https://doi.org/10.3102/0034654314564881

van Dooren, M., de Vries, J. J., & Janssen, J. H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiol Behav*, *106*(2), 298–304. https://doi.org/10.1016/j.physbeh.2012.01.020

Veenman, M. V. (2016). Learning to self-monitor and self-regulate. In *Handbook of research on learning and instruction* (pp. 249–273). Routledge.

Venkatesh, Morris, Davis, & Davis. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, *27*(3), 425. https://doi.org/10.2307/30036540

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Wauters, K., Desmet, P., & Van den Noortgate, W. (2010). Adaptive item-based learning environments based on the item response theory: Possibilities and challenges: Adaptive ITSs based on IRT. *Journal of Computer Assisted Learning*, *26*(6), 549–562. https://doi.org/10.1111/j.1365-2729.2010.00368.x

Winne, P. H. (2010). Improving Measurements of Self-Regulated Learning. *Educational Psychologist*, *45*(4), 267–276. https://doi.org/10.1080/00461520.2010.517150

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In *Handbook of Self-Regulation* (pp. 531–566). Elsevier.

Wohlfart, O., & Wagner, I. (2022). "DigitalPakt Schule 2019 bis 2024" – Analyse der strukturellen Digitalisierung des Bildungssystems in Deutschland. *Praxis Forschung Lehrer*innen Bildung. Zeitschrift für Schul- und Professionsentwicklung.*, *4*(1), Article 1. https://doi.org/10.11576/pflb-5973

Wong, J., Khalil, M., Baars, M., De Koning, B. B., & Paas, F. (2019). Exploring sequences of learner activities in relation to self-regulated learning in a massive open online course. *Computers & Education*, *140*, 103595.

https://doi.org/10.1016/j.compedu.2019.103595

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. *Handbook of Metacognition in Education* (pp. 311–328). Routledge.

Appendix A – Manuscript A (Study 1)

What happens to your body during learning with computer-based environments? Exploring negative academic emotions using psychophysiological measurements

Abstract

This explorative study aims to examine if electrodermal activity (EDA) and heart rate (HR) are appropriate measures for identifying and monitoring academic emotions during learning in computer-based learning environments (CBLEs). Understanding learners' emotions while using CBLEs, allows improving the design of CBLEs. Therefore, we collected EDA, HR, and self-report data from 32 participants to measure academic emotions during learning with CBLEs in a laboratory setting. We induced negative academic emotions during learning using harmful connotated learning content about animal welfare. In a pre-post design, participants reported their emotional state before and after learning. We collated the self-reports with the EDA and HR curves to identify the emotional change in real-time. We prepared the data for repeated measurement analyses and group differences (high-, middle-, low learning performance; bored vs. not bored participants). Negative academic emotions were detected in increased EDA and HR. EDA turned out to be an indicator of learning performance. Boredom manifested in HR decrease. Findings show that EDA and HR are appropriate tools to measure academic emotions. We want to show the importance of real-time measures for learning and the efficiency of EDA and HR measures. It is worth considering EDA as a predictor for learning success and implementing EDA and HR measurements in CBLEs. However, more research is needed to clarify the role of HR in the context of learning performance.

Keywords: academic emotions, electrodermal activity, heart rate, computer-based learning environments, learning processes

There is no doubt that emotions influence our learning behavior and outcome. When we are in a good mood, we learn more successfully (Arguel et al., 2017; Duffy et al., 2018; Loderer et al., 2020). This widely replicated insight shows that emotional states significantly impact learning performance (for a review, see Panadero, 2017; Loderer et al., 2020). Therefore, we considered it essential to explore emotions in the context of learning further.

Since we investigated computer-based learning environments (CBLEs), a specific set of emotions comes to the fore: emotions occurring in educational settings (e.g., studying at home, taking an exam, or being in class) are defined as *academic emotions* and are directly bound to learning and achievement. The most-reported academic emotions are anxiety, enjoyment, hope, pride, relief, anger, boredom, and shame (e.g., Duffy et al., 2018; Järvenoja et al., 2017; Loderer et al., 2020; Pekrun et al., 2002). Academic emotions are mainly evaluated post facto using self-report data (e.g., Boekaerts, 1999; Eteläpelto et al., 2018; Magno, 2011; Pekrun et al., 2011, 2017; Vermeer et al., 2000). However, a notable drawback of self-reports is that emotional states must be experienced consciously to report on them. Collecting post facto and self-report data reveals subjective responses about past events, which can cause measurement errors (e.g., Arguel et al., 2017; Laarni et al., 2015; Slater, 2002). Nevertheless, self-report data is a crucial and meaningful tool to gather subjective experiences, but it is limited according to an objective and implicit exploration of emotional processes during learning.

A promising approach to evaluate learning processes besides self-reports is "on-the-fly" measures stated by Winne and Perry (2000). Also, Järvelä and colleagues (2019) showed that analyzing real-time data is fruitful. They explored self-regulated learning by using qualitative content analyses, facial expressions, and psychophysiological measurements (i.e., electrodermal activity [EDA] and heart rate [HR]) in a collaborative learning setting. Confusion, for example, was detected based on a simultaneous increase in EDA, negative facial expressions, and a complimentary content analysis (Järvelä et al., 2019).

Our research goal is to provide deeper insights into learning (i.e., progression of the learning process besides self-reports, see chapter 1.4.) and explore the psychophysiological appearance of academic emotions. Based on the findings mentioned above and to balance the mentioned limitations of self-reports, the present study relied on psychophysiological measurements (i.e., EDA and HR) to examine academic emotions in CBLEs. Because changes in physiological behavior can have multiple reasons (see chapter 1.3.), we eliminated as many confounding factors (e.g., the impact of social interactions in collaborative learning settings or movement artifacts) as possible by using a straightforward laboratory set-up. More precisely, we explored if specific physiological response patterns can be found, which indicate the current emotional state of the learner. Moreover, we seek to determine if physiological behavior can be a sufficient indicator for learning performance. Furthermore, we analyzed the change of academic emotions before and after stimulus presentation and whether this progress is evident in psychophysiological data.

Theoretical framework

The Dual Processing Self-Regulating Model

The *Dual Processing Self-Regulating Model* from Boekaerts (2011) describes the essential role of emotions in learning. Boekaerts (2011) claimed that emotional states guide the learner's behavior onto one of two possible pathways. She proposed a *well-being* and a *growth pathway* as self-regulatory strategies, depending on how the task is assessed. Tasks that do not fit the current mental model trigger negative emotional states, which are detrimental for knowledge increase, leading the learner to take the well-being pathway. Tasks that correspond with the learner's goals cause positive emotional states and thus open the growth pathway, resulting in knowledge increase. Measuring learners' emotional states can therefore propose a statement about learning success.

Furthermore, it is possible to switch from one pathway to the other. If learners are on the growth pathway and detect indicators for failing, they shift to the well-being pathway (Boekaerts, 2011). Determining this emotional shift in real-time enables immediate support and therefore guides the learner back on the growth pathway (see Arguel et al., 2017; D'Mello & Graesser, 2014). We want to find an appropriate "on-the-fly" measure that can identify negative emotional states during learning with CBLEs, as a step towards the primary goal of guiding and keeping the learner on the growth pathway.

Academic emotions

Given that emotions are concomitants of learning, it is necessary to differentiate these academic emotions specifically (Pekrun & Stephens, 2012). Academic emotions, which can be seen in Table 1, are related to achievement, classroom settings, and learning. They are bound to success and failure, but also to the process of learning itself (Goetz &

Hall, 2013; Pekrun et al., 2002, 2017). Multiple research approaches address academic emotions (e.g., confusion: D'Mello et al., 2014; boredom: Goetz & Hall, 2013; Pekrun, 2006; Pekrun et al., 2002). The underlying concept of this work is the *Three-Dimensional Taxonomy of Academic Achievement Emotions* from Pekrun (2006), which classifies academic emotions in three dimensions: their valence (positive or negative), activation (activating or deactivating), and object focus (activity or outcome; see Table 1). Enjoyment, for example, is, according to Pekrun (2006), a positive and activating academic emotion, during an activity (e.g., studying). In comparison, sadness is defined as negative and deactivating academic emotions triggered by pro- or retrospective failure (e.g., upcoming or past exams).

In the psychophysiological literature, the term "arousal" is more common than activation (e.g., Berntson et al., 2017; Lang et al., 2009; Levenson et al., 2017; Potter & Bolls, 2012). To have consistent terminology in this article, we refer to the term "activation".

Table 1
A Three-Dimensional Taxonomy of Academic Achievement Emotions

	Positive[a]		Negative[b]	
Object Focus	Activating	Deactivating	Activating	Deactivating
Activity	Enjoyment	Relaxation	Anger	Boredom
Outcome	Joy Hope Pride Gratitude	Contentment Relief	Anxiety Shame Anger	Sadness Hopelessness Disappointment

Note. Academic Achievement Emotions categorized into three dimensions valence, activation, and object focus. [a]Positive = pleasant emotion. [b]Negative = unpleasant emotion (based on Pekrun & Stephens, 2012, p. 4).

Negative academic emotions usually trigger task-irrelevant thoughts and decrease the resources required for the task. Therefore, learning performance may decline if a learning goal seems unachievable due to prevalent negative academic emotions. However, negative activating academic emotions can also cause intense motivation to prevent failure, resulting in solving the task and increasing learning performance (Pekrun & Stephens, 2012). The shift from detrimental and conducive emotional states is also supported by Boekaerts' *Dual Processing Self-Regulating Model* (2011; see chapter 1.1.), where learners switch from the well-being pathway to the growth pathway. Depending on the learner's assessment and the apparent solvability of a task, emotional states can change, and even knowledge can increase despite experiencing negative emotions during learning (Boekaerts, 2011).

Furthermore, task difficulty can affect academic emotions due to cognitive incongruity (Pekrun & Stephens, 2012). If the task seems too tricky or non-solvable, negative academic emotions are triggered, resulting in low learning performance (Baker et al., 2010; D'Mello & Graesser, 2014). Otherwise, positive academic emotions arise if a learning task can be solved, leading to high learning performance (Kang et al., 2008; Pekrun & Stephens, 2012).

In the present study, we decided to focus on negative activating academic emotions to reduce complexity. Besides, it is more valuable to properly understand the physiological appearance of negative academic emotions and cope with them to promote learning. We are interested in whether learners show an increase in knowledge despite the task causing negative academic emotions, or say it with Boekaerts' approach if there is an increase in learning, a shift from the well-being to the growth pathway has happened.

Psychophysiological measurements for academic emotions

Psychophysiological measures (e.g., EDA, electromyography, eye-tracking, or electrical activity of heart and brain) are well-elaborated to index cognitive tasks and emotional states (see Berntson et al., 2017; Dawson et al., 2017; Levenson et al., 2017). Psychophysiological measurements aim to conclude from physiological reactions to psychological processes (e.g., emotions or attention; Pinel & Pauli, 2012). Here, the essential statement is that physiological processes are intertwined with human behavior (Cacioppo et al., 2017). Based on psychophysiological data, conclusions concerning emotional processes can be drawn. Psychological conditions cannot be associated with a separate isolated physiological reaction. The complex reaction pattern must always be considered (Cacioppo & Tassinary, 1990). For example, an electrodermal reaction can indicate an arousing situation or a deep breath. Both situations show the same result - an increase in the electrodermal curve - but they are very different in their respective meaning. Therefore, there is no one-to-one relation between a single physiological response (e.g., an increase in EDA or HR deceleration) and a specific emotion (e.g., frustration). For example, an increase in EDA cannot identify frustration, and frustration does not express solely in changing EDA. Adding HR as a measure for valence can specify the increase in EDA since negative emotions express in HR decrease (see chapter

1.3.1. for EDA and 1.3.2. for HR). Therefore, the psychophysiological pattern composed of EDA and HR curves must be considered to identify emotional states. The attribution from physiological response patterns to actual psychological meaning requires an accurate experimental design, appropriate data analyses, and interpretation (Cacioppo et al., 2017).

Since we see emotions as a two-dimensional model, both, valence and activation must be examined to capture emotions comprehensively. Then, merging EDA and HR data reveals a physiological pattern, which can identify emotional states (e.g., Barrett & Russell, 1999; Eteläpelto et al., 2018; Larsen & Diener, 1992; Levenson et al., 2017). Furthermore, only the valence can declare if the emotion is positive or negative, which is crucial for successful learning. We chose EDA and HR since these are easily measurable, non-invasive, sensitive to psychological states, and well-elaborated (see chapters 1.3.1. and 1.3.2.). Based on established research about psychophysiological measurements, we used EDA to capture the activation and HR to measure the valence of academic emotions. We do not further address the third dimension "object focus" because it refers to whether the emotional state is seen as activity or outcome (see Table 1), which is not relevant for our purpose.

Electrodermal activity

A standard psychophysiological measurement in many different research areas is EDA (e.g., attention, information processing, and emotion). Its popularity is the simple measurability and the sensitivity to many psychological states and processes (Dawson et al., 2017). EDA changes are associated with emotional activation, emotionally arousing thoughts or events, which induce an increase of electrical conductivity of the skin (Bradley, 2009). The EDA is solely controlled by the sympathetic nervous system (SNS) and, therefore, a direct reflection of activation (details see chapter 1.3.2.; Dawson et al., 2017; Lang et al., 2009). The interpretation of EDA changes depends on the stimulus material and the surroundings (Dawson et al., 2017). For example, an increase in EDA in an emotional surrounding can be interpreted as increased emotional activation. When somebody gets frightened, the increase in EDA can be traced back to the occurring attentional shift towards the unexpected stimulus (Bradley, 2009). Therefore, the more controlled a laboratory setting is, the more reliable is the interpretation of a change in EDA (Dawson et al., 2017). Moreover, having more than one measure (e.g., HR and self-reports) leads to a more accurate reconstruction of the learner's psychological state (Lang, 2014).

The most used method of recording EDA are skin conductance level (SCL) and skin conductance response (SCR), both measured in microSiemens (µS). The tonic SCL measures the level of skin conductance in a particular situation and ranges from two to 20 µS. The phasic SCR shows temporary fast changes in the skin conductance caused by discrete events and ranges from one to five µS (Dawson et al., 2017).

Heart rate

Besides the primary function of pumping blood through the body, the heart also reveals information about emotion, attention, activation, and information processing (Berntson et al., 2017; Lang et al., 2009; Potter & Bolls, 2012). HR is, like EDA, easily measurable, non-invasive, and associated with many different psychological states. The HR shows the frequency of a cardiac cycle and is measured in beats per minute (bpm; Berntson et al., 2017). The most promising measurement is an inter-beat interval (IBI). Here, the time between two peaks of the cardiac cycle is tracked. The most prominent peak of the cardiac cycle is the R-spike. The time between two R-spikes is called RR-interval (Potter & Bolls, 2012).

Fluctuations in the HR can tell if a stimulus is pleasant or unpleasant, meaning HR is sensitive for measuring valence (Greenwald et al., 1989). Pictural stimuli (everyday objects or exciting scenes), which were assessed as pleasant (e.g., a beautiful landscape or erotic pictures), lead to HR acceleration, and pictural stimuli, assessed as unpleasant (e.g., dirty laundry or mutilated bodies), cause HR deceleration (Ijsselsteijn et al., 2000; Lang et al., 1993, 1997; Palomba et al., 1997). The valence of the pictural stimuli (pleasant or unpleasant) was evaluated and standardized by the International Affective Picture System, which can be used to explore emotion and attention (Lang et al., 1997).

Nevertheless, it is reasonable to assume that activating emotions lead to HR acceleration and deactivating emotions to HR deceleration. However, this relation does not necessarily persist based on the mechanics of the autonomic nervous system, which regulates HR and EDA. The link between activation and valence regarding the HR underlies the dual control of the heart. Its pace is regulated by both autonomic nervous branches, the parasympathetic nervous system (PNS), and the SNS (Berntson et al., 2017; Lang et al., 2009; Levenson et al., 2017). Both systems influence how fast the heart beats, depending on which system is activated. The activation of the PNS leads to HR deceleration, which is associated with attention and cognitive effort (Lang et al., 2009). The activation of the SNS results in HR acceleration, which is related to emotional activation (Lang, 1994). Therefore, HR can be a measure of valence but also activation. Nevertheless, since the PNS is faster and more dominant than the SNS, the activation of the SNS must be potent to overcome the parasympathetic activation (Shaffer & Ginsberg, 2017). A parameter to determine which system is activated is the heart rate variability (HRV), measured by spectral analyses (Berntson et al., 2017; Shaffer & Ginsberg, 2017).

Purpose of the study and research questions

When we consciously experience emotions like love, happiness, anxiety, or distress, we feel our physiological reactions (e.g., faster heartbeat or sweaty hands). However, unconscious emotional states, especially in the context of learning, equally impact our physiological behavior and are thus detectable in psychophysiological curves. Furthermore, psychophysiology allows visualizing emotional processes in real-time (see chapter 1.3.).

Various studies have explored emotions in CBLEs and collaborative learning settings in a diverse manner (for a review, see Loderer et al., 2020). However, psychophysiological assessments of academic emotions in educational psychology are underutilized (Pekrun & Stephens, 2012). The present study wants to address this issue and get a unified and clear perspective on academic emotions, CBLEs, and self-reports. Moreover, we captured the valence and activation of academic emotions separately to give a detailed statement about the psychophysiological appearance of academic emotions. It was realized with a simple study design in a laboratory set-up (see Fig. 1) that eliminates potential external influencing factors (e.g., big-fish-little-pond effect; Preckel et al., 2008). The learning setting was designed to evoke negative emotions and guide the learner onto the well-being pathway. This process aims to be made physiologically detectable. Due to the lack of literature, the present work's research question and data analyses were primarily exploratory.

Since psychophysiological reactions unfold over time, they are an adequate measurement for academic emotions, which also occur over time. Self-reports give information about an emotional pre- and post-state of the learner – but they cannot provide details about the progression or reasons for the emergence of emotions. The exploratory research question (RQ) and hypotheses are structured top-down with the broad RQ at the top and the detailed hypotheses at the bottom. The derived RQ targets whether physiological behavior reveals more information about academic emotions and learning:

- Can psychophysiological measurements provide deeper insights into learning processes?

The explorative character of the RQ allows space for different data analyses and approaches. The term "deeper insights" implies getting information about the ongoing learning process (psychophysiological data) rather than solely having information about the current state of knowledge (self-reports). Moreover, the cause, emergence, and physiological progression of academic emotions provide insights into learning behavior. We formulated detailed hypotheses to follow the top-down approach, referring to negative academic emotions and their physiological indicators.

The hypotheses target specific data analyses to find distinct physiological patterns and thus indicators of academic emotions. We state that patterns in EDA and HR indicate negative academic emotions. To meet the requirements of the two-dimensional model of emotions, we formulate a particular hypothesis for each dimension. Valence is captured by HR, and EDA captures activation.

Learning requires attention and information processing, which activates the PNS. In the psychophysiological context, this implies that the HR decreases. Moreover, the designed learning environment (see chapter 2.3.) included unpleasant stimuli, leading to HR decrease (see chapter 1.3.2.). Therefore, we state:

- Negative activating academic emotions cause HR deceleration over time (H1).

Emotional activating situations cause an increase in EDA (see chapter 1.3.1.). We want to show that this condition transfers to learning (i.e., academic emotions). The learning materials (see chapter 2.3.) induced negative activating academic emotions. Thus, we state:

- Negative activating academic emotions cause increasing EDA over time (H2).

To associate learning, HR, and EDA, we formulated the third hypothesis. Task difficulty, analyzed using learning performance, has an impact on academic emotions (see chapter 1.2.), which can be measured by changes in EDA and HR:

- Depending on the learning performance (high vs. low), overall HR and EDA differ (H3).

In conclusion, the *Dual-Processing Self-Regulating Model* (Boekaerts, 2011) shows that emotions have a crucial impact on learning (see chapter 1.1.). Since learners cannot always detect detrimental academic emotions, learning success can be affected negatively. We want to show an approach, which makes academic emotions measurable in real-time so that learners can be supported immediately. EDA and HR provide a fruitful measurement for emotions (see chapter 1.3.). Based on the *Three-Dimensional Taxonomy of Academic Achievement Emotions*, we aim to measure both, valence and activation to distinguish between detrimental and beneficial academic emotions (the third dimension "object focus" has no further relevance for our approach, see chapter 1.3.). Anger and enjoyment, for example, are both activating but different in their valence. Only if both dimensions are measured, detrimental (e.g., anger) and beneficial (e.g., enjoyment) can be discriminated, and the learner can be supported accurately.

Method

Participants

Acquisition of participants was realized via a web-based online recruitment system *ORSEE* (Greiner, 2015). Participants were students and employees from the Technical University of Munich (N = 32; 21 females; M_{age} = 27.82, SD = 2.45). The inclusion criterion was being fluent in German to understand the stimulus material perfectly. We excluded one participant because of insufficient concentration and individual data channels with poor psychophysiological recordings. This results in different sample sizes for self-reports: n = 31 (20 females), HR: n = 28 (18 females), EDA: n = 27 (16 females). Despite the small sample size, a sufficient test power (β = .80) according to an a-priori analysis (α = .05) can be achieved, which suggested 30 participants for mildly correlated repeated measures (r = .20) with a minimum of 16 number of measurements without baseline (Faul et al., 2009). Based on the mixed findings on whether emotions can be discriminated by indicating EDA and HR, we assume a medium effect size of f = .25 (Berntson et al., 2017; Boucsein, 2012; Levenson et al., 2017). Since we want to consider as much data as possible, we focused on the first 17 data points (incl. baseline), where all participants are included.

Measures

Self-reports

We used the German versions of the Positive and Negative Affect Schedule (PANAS, Krohne et al., 1996; $\alpha \geq .84$; 5-point Likert-scale) and the seven-item short version of the Epistemically-Related Emotion Scale (EES-D, Pekrun et al., 2017; $\alpha \geq .76$; 5-point Likert-scale) in a pre-post design to measure the change of perceived emotional states after learning. We combined PANAS and EES-D because PANAS covers the overall emotional state (Krohne et al., 1996), and the EES-D refers to emotions accompanied by cognitive activities and knowledge generation (Pekrun et al., 2017; Pekrun & Stephens, 2012). Both questionnaires measure emotional activation and valence subjectively and are collated to EDA and HR as an objective measure for activation and emotional valence. The Academic Emotions Questionnaire (AEQ, Titz, 2001; $\alpha \geq .84$; 5-point Likert-scale) was only included in the posttest to retrieve information about the emotional experience of the previous learning situation. The AEQ consists of class-, learning-, and test-related emotion scales, which can be applied separately. Since we focus on the learning situation itself, we chose the learning-related emotion scale, which includes eight subscales (enjoyment, hope, pride, anger, anxiety, shame, hopelessness, boredom). Each item of the AEQ refers either to emotional experiences before, during, or after learning. To not overwhelm the participants, we used the 45 items of the AEQ, which gathered experiences during learning. The AEQ does not primarily refer to the valence or activation of emotions but mainly to the emotional evaluation of learning. Moreover, a short-form of a resilience scale (RS-13, Leppert et al., 2008; $\alpha = .69$; 7-point Likert scale) was used before learning to determine possible correlations with emotional states and physiological behavior (prototypical items of the mentioned scales can be seen in Table 14 in the supplementary material). Learning performance was measured using a self-designed questionnaire with 10 multiple-choice items and one open question immediately before (prior knowledge) and after the learning session. (e.g., "*Conventional housing conditions for animals violate animal welfare laws. Why?*" or "*What is animal-turn?*" followed by four answer options). The score of the prior knowledge was subtracted from the score, which participants achieved after learning and is used to represent learning performance. To minimize guessing, participants always had the chance to mark "*I don't know*". The open question queried a correct abbreviation for a technical term and was rated with one point for the correct spelling. Regarding the multiple-choice items, participants scored for marking the correct answer and not marking the incorrect answer with one point each, resulting in a maximum score of 33. All items refer to the content of the learning material, which measures knowledge increase after learning.

Consequently, the pretest contained PANAS and EES-D measuring the current emotional state, RS-13 gathering an unbiased value of resilience, and the content-related questionnaire testing prior knowledge. The posttest included PANAS and EES-D gaining the perceived change of emotional states, the content-related questionnaire measuring knowledge increase, and AEQ gathering the emotional experience of the previous learning situation. All scales and descriptive statistics for the present study can be seen in Table 2.

Table 2
Results for the Self-Report Measures for Negative Emotions and Scale-Reliability

Measure		No. of items	Min.	Max.	*M*	*SD*	Cronbach's α
PANAS[c]	negative affect	10	1.07[a]	1.94[a]	1.30[a]	0.31[a]	.747[a]
			1.39[b]	3.20[b]	2.36[b]	0.73[b]	.877[b]
EES-D[c]	confused, anxious, frustrated, bored	4	1.36[a]	1.52[a]	1.4[a]	0.49[a]	.632[a]
			1.39[b]	2.36[b]	1.74[b]	0.59[b]	.622[b]
AEQ[c]	anger, anxiety, shame, hopelessness, boredom	32	1.74[b]	2.52[b]	2.12[b]	0.64[b]	.868[b]
RS		13	4.68[a]	5.90[a]	5.23[a]	0.35[a]	.687[a]
Learning performance[d]		11	6[a]	18[a]	11.7[a]	3.08[a]	-
			17[b]	26[b]	21.7[b]	2.48[b]	-

Note. N = 31.
[a]pretest, [b]posttest, [c]itemized by valence, [d]maximum score = 33.
PANAS = Positive And Negative Affect Schedule; EES-D = Epistemically-Related Emotion Scale; AEQ = Academic Emotions Questionnaire; RS = Resilience Scale.

Psychophysiological data

We used the BIOPAC MP36 system and the *Biopac Student Lab 4.1* software to record and process physiological data sampled with a 1 kHz rate. We sampled at a high rate to have valid data after smoothing and removing artifacts (see Boucsein et al., 2012). For proper measurements, we used the SS57L lead set and disposable snap Ag/AgCl pre-gelled electrodes for EDA and the fully shielded cable SS2LB with Ag/AgCl disposable snap pre-gelled electrodes EL501 for HR. From raw HR data, RR-intervals were derived in real-time for later analyses. Raw EDA data was treated with a 1 Hz FIR low-pass filter, and phasic data was derived from the tonic curve using a 0.05 Hz IIR high-pass filter. Artifacts were treated additionally with smoothing routines or interpolation methods. Furthermore, the baseline mean was subtracted from the curves to obtain standardized values and comparable data among all participants. The resulting channels with physiological data were resampled with 100 Hz and exported as text and excel files for further analyses.

The entire sequence of the study, that is, stimulus material, the participant's screen, and the recording of the participants - especially the placements of the electrodes, was recorded with *iMotions* version 8.1.

Learning environment

We chose unpleasant stimuli as learning material to direct the participants on the well-being pathway and induce negative emotional states. The video is an actual report made by the public-sector broadcaster. The video consists of recordings made from animal welfarists in conventional pig farms and scenes of Germany's political discussion about animal welfare. It starts with dramatic music and a voiceover, who reports about the illegally recorded scenes from pigsties, which were used to call attention to the mischief in conventional pig farming, triggering scare (see Fig. 1, on the left). Following scenes from a political event, the federal minister of Food and Agriculture (Germany) speaks about the danger that animal welfarists pose when recording illegally and that animals are protected by law. These scenes evoke an imbalance between reality and politics. Subsequently, the legal basis of conventional pig farming is presented. The conclusion is that many pig farms and the welfare of animals were not appropriately controlled, which activates anger. Then the illegally recorded scenes from pigsties continue, leading to sadness and distress. The voiceover continuously reports about the legal basis, the political discussion, and the animal protection act. The following scenes show how piglets were killed by an employee, which triggers distress and anger. The video concludes that the violation of the animal protection act is not punished sufficiently, resulting in frustration. Overall, the video induces severe negative emotions.

Afterward, the participants had to read a challenging scientific paper from Bruhn and Wollenteit (2018) about the detailed legal basis of the animal protection act and regulations. The text includes a lot of paragraphs and laws, which makes it difficult to read and understand (see Fig. 1, on the right). Because the participants were told to understand recall as much information as possible, the task gets more difficult or even unsolvable, which should maintain the negative mood and lead to frustration and eventual boredom. The overall learning environment should affect the ongoing task appraisal in an emotionally negative manner, leading to perceived insolubility of the task. Therefore, a shift to the well-being pathway, indicated by changing psychophysiological behavior.

We pretested the learning material separately to ensure that it triggers negative emotions ($N = 5$). These pretests show that both stimuli evoke negative emotions ($p < .05$ for distressed, scared, hostile, upset, ashamed; detailed t-tests see Table 5 in the supplementary material).

Fig. 1
Screenshots of the Learning Material

Note. Screenshot of the video on the left, an excerpt of the text on the right.

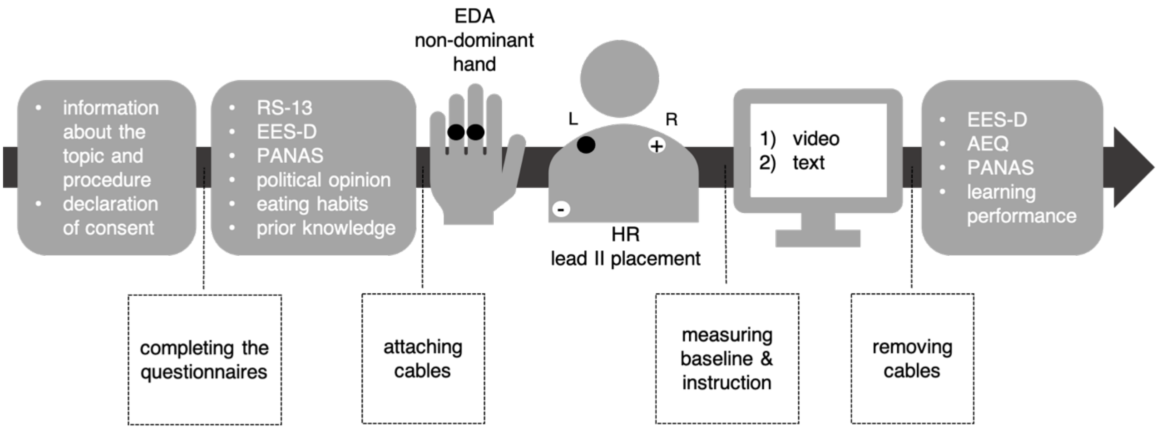
Procedure

Initially, we informed participants about the procedure of the study and psychophysiological data collection. We only shared the topic but no hypotheses or research interests. Then, participants had to sign a declaration of consent. Before the learning session started, participants answered questionnaires about resilience (RS-13), epistemic emotions (EES-D), current emotional states (PANAS), their political opinion about pig farming, eating habits, and prior knowledge about the topic to generate the learning performance score. During a rest period of five minutes, electrodes for the psychophysiological measurements were applied, which ensures an even hydration between the electrode, gel, and skin.

Moreover, the participants could get used to the laboratory set-up while a baseline was measured. Two electrodes were applied to the palmar proximal phalanges of the middle and ring finger of the non-dominant hand to record EDA. To collect HR data, we attached three electrodes according to the lead-II placement and the Einthoven Triangle to the upper body (two electrodes under the collarbone, one electrode on the left side of the ribcage, see Fig. 1). The learning session consisted of the six-minute video followed by the scientific text described in chapter 2.3. that started automatically after the baseline measurement using *iMotions* (version 8.1). We instructed the subjects to pay attention to the content and memorize as much information as possible immediately before the learning session. When the participants finished reading, cables and electrodes were removed. Afterward, information about the level of knowledge (learning performance), epistemic (EES-D), and academic (AEQ) emotions and current emotional state (PANAS) were gathered, and participants were informed about the research questions. The entire study lasted about one hour and took place in a laboratory of the Technical University of Munich.

Fig. 2

The Study Design, Including Every Step of the Procedure, all Instruments, and Placement of the Electrodes in Chronological Order From Left to Right



Data processing

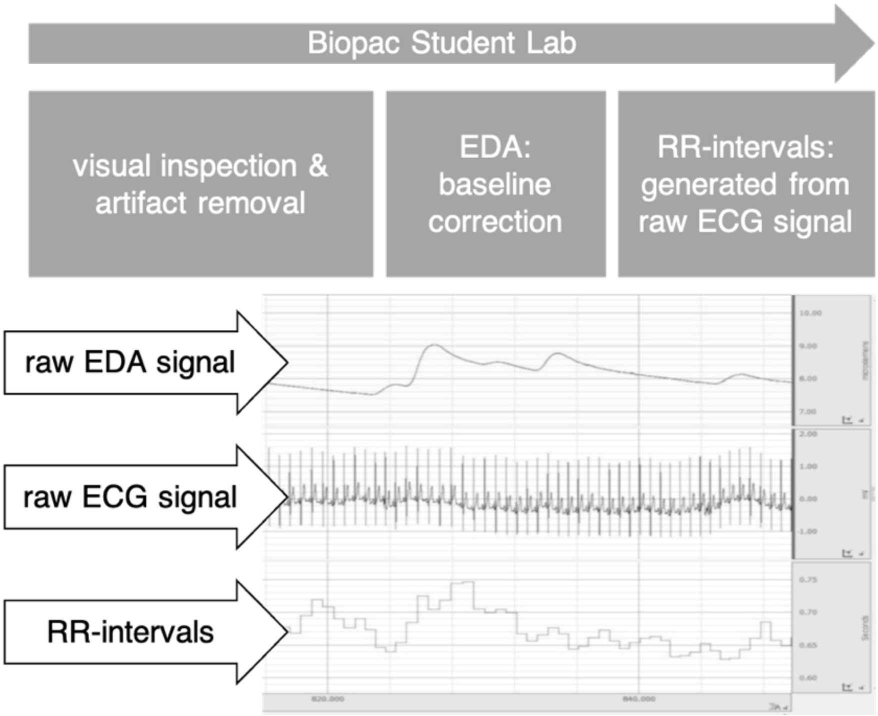
All self-reports were collected using the online survey tool *SoSci Survey* and analyzed using *SPSS Statistics 26* (IBM Corp., 2020) and *JASP* (JASP Team, 2020).

The following treatments were recommended by the software creators (Sjak-Shie, 2019) and carried out in scientifically replicated standard procedures (see Boucsein et al., 2012; Cacioppo et al., 2017; Potter & Bolls, 2012).

First, each data channel was checked visually for measurement errors or artifacts, and if necessary, smoothing or artifact removal procedures were used. The EDA signal was baseline corrected. The baseline correction is necessary because EDA can vary widely between and within participants (2 – 20 μS; see chapter 1.3.1.) We subtracted the baseline, which was measured before the learning session (see Fig. 2) for each participant individually to generate comparable curves. The RR-intervals were generated in real-time from the raw electrocardiogram (ECG) using a standard procedure provided by *Biopac Student Lab*. Each step of data processing in the *Biopac Student Lab* and a screenshot of data recordings can be seen in Fig. 3.

Fig. 3

Chronological Steps of Data Processing in the Biopac Student Lab including a Screenshot of Data Recordings



Note. The displayed data stems from one of our participants.

To analyze HRV, we used the MATLAB-based application *PhysioDataToolbox* version 0.5 (Sjak-Shie, 2019). Therefore, the raw ECG signal was extracted from the *Biopac Student Lab*. The ECG signal analyzer treated the raw ECG data with a 1 Hz high-pass filter and a 50 Hz low-pass filter. To detect and count R-spikes, the minimum value of 0.38 millivolt and the minimum distance of 0.3 seconds between R-spikes must be fulfilled. Peaks below or above these values were not classified as R-spikes (see Fig. 4a on the left). Then, IBIs were derived from the detected R-spikes. A minimum value of 0.4 seconds and a maximum value of 1.3 seconds between the R-spikes must be fulfilled to be classified as IBI. IBIs with lower or higher values than these parameters were automatically rejected (see Fig. 4a on the right). The HRV analyzer used these generated IBIs and resampled them with a 4 Hz frequency. A spectral analysis was carried out to get information about which frequency components account for the variability of the heartbeat. Therefore, a very low (0.0033 Hz & 0.04 Hz), low (0.04 Hz & 0.15 Hz) and high (0.15 Hz & 0.4 Hz) filter power band were calculated. The resulting curves reveal whether the PNS (high-frequency) or the SNS (low-frequency) controls the heartbeat (see Fig. 4), which allows a proper interpretation of the HR data and their psychological meaning. The most descriptive output was the percentage distribution of each filter power band, and thus, if PNS or SNS controls the HR. The very low filter power band stands for thermoregulation, which is not relevant in our case.

After processing participants separately, we integrated all data in one file and visually lapped every data channel to identify outliers or abnormal curves between participants.

Finally, we exported all psychophysiological data in one excel-file for statistical analyses. We used the generated HRV data from *PhysioDataToolbox* and the data processed in *Biopac Student Lab* to analyze EDA, HR, and HRV data statistically.

We used two different methodical approaches. First, we prepared the data for repeated measurement analyses and group differences. Since we do not have specific areas or a stimulus onset but are interested in the progression of the curves over time, we averaged each data channel per minute, resulting in at least 17 (incl. baseline) values per participant (Min = 17, Max = 39; for HR: M = 27.0, SD = 5.54; for EDA: M = 27.1, SD = 5.64). These data segments were recommended by the software creators (Sjak-Shie, 2019). To avoid confusion: increasing HR represents decreasing RR-intervals.

Fig. 4

Illustration of Generating the Heart Rate Variability in the PhysioData Toolbox

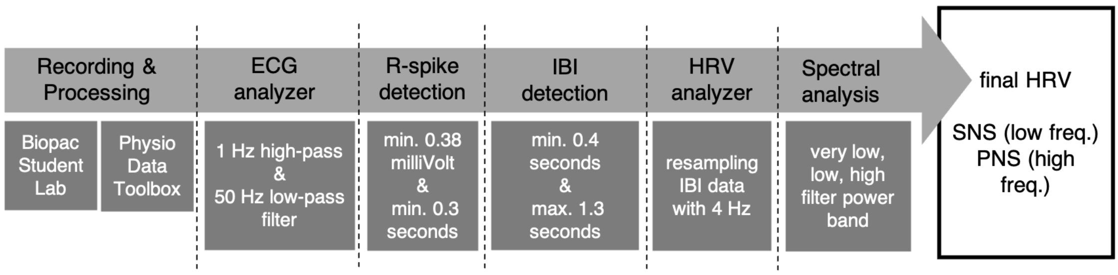
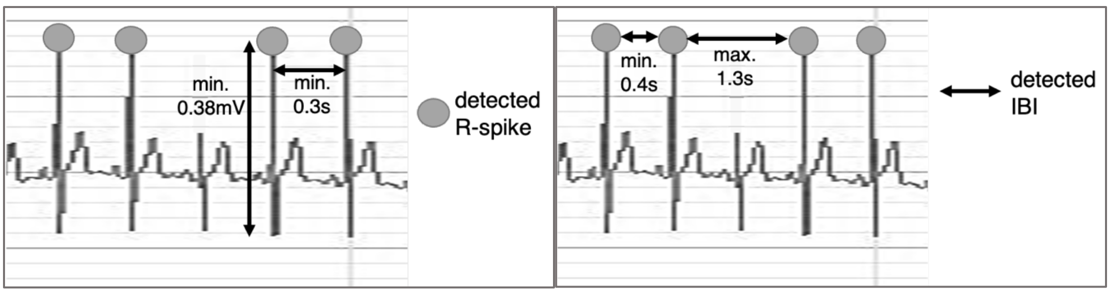


Fig. 4a

Illustration of how R-spikes (on the Left) and Inter-Beat-Intervals (on the Right) Were Detected



Note. The displayed data stems from one of our participants.

To test H1 and H2, we conducted an ANOVA with repeated measurements to analyze if, when, or where psychophysiological curves differ. Therefore, we can explore how the curves progress over time. Most important when analyzing psychophysiological data is the visual inspection. Thereby, artifacts can be detected and removed easily. Afterward, an ANOVA with repeated measurements can be used as trend analysis. Here, the shape of the curves can be described. If a linear trend can be shown, the curves follow a linear progression. If the curves would fluctuate intensively, quadratic or cubic curves could be found, which is not to be expected in our case. Moreover, we conducted HRV analyses to determine which nervous system (i.e., PNS or SNS) controls the HR (see chapter 1.3.2.).

To test H3, we performed simple linear regression analyses, with EDA or HR as the predictor and learning performance as the dependent variable. Additionally, we performed a One-Way ANOVA to look for group differences in learning performance (high vs. middle vs. low).

It was noticeable that some participants were less bored after the learning phase than before during data collection. Therefore, we compared the psychophysiological curves of these participants to find patterns.

Findings

Self-reports

The presented results regarding emotional states stem from the EES-D, PANAS, and AEQ questionnaires.

All participants reported a significant increase in negative emotional states after learning (e.g., frustrated, distressed, scared, upset), indicating a negative appraisal of the task. However not significant, an unexpected tendency to a decrease in self-reported boredom after learning can be shown (see Table 3), which is in line with the verbal feedback from the participants. They expressed interest in the topic and wanted to receive more information.

The learning performance was significantly higher in the posttest (see Table 3). Learning performance scores were normally distributed (Shapiro-Wilk test $p = .95$). Due to technical problems, poor psychophysiological data, or artifacts, the sample size varied. Detailed descriptive information can be found in the supplementary material.

Table 3
Self-report values for negative academic emotions and learning performance

	M_{pre}	SD_{pre}	M_{post}	SD_{post}	$t(30)$	p	Cohen's d
Frustrated	1.42	0.67	2.35	1.14	4.21	< .001	0.76
Distressed	1.55	0.85	3.00	1.07	6.86	< .001	1.23
Scared	1.06	0.25	3.13	1.38	8.58	< .001	1.54
Upset	1.13	0.34	3.19	1.20	8.92	< .001	1.60
Bored	1.52	0.77	1.39	0.56	-0.94	.35	-0.17
Learning performance	11.7	3.08	21.7	2.48	13.2	< .001	2.37

Note. N = 31.

Psychophysiological data and learning performance

In the following sections, additionally, to test the hypotheses, exploratory analyses were carried out.

Simple linear regression analyses were used to examine whether psychophysiological behavior can predict learning performance. EDA data (i.e., the average skin conductance level) was used as a predictor and learning performance (i.e., difference score) as a dependent variable. The model showed a R^2 of .27 (adjusted R^2 = .24, $F(1, 26) = 9.62$, $p = .005$, $\beta = -.52$), which indicated, according to Cohen (1988) a high goodness-of-fit. EDA was therefore a significant predictor for learning performance, $t(27) = -3.10$, $p = .005$. Regression analyses for HR data (i.e., average HR in bpm) did not show a convenient fit ($F(1, 29) = 0.38$, $p = .54$).

As exploratory analyses an ANOVA with repeated measurements was conducted using 60-second-slices for EDA and HR (see chapter 2.5.). EDA and HR curves followed a significant linear trend. EDA ($F(1, 26) = 10.4$, $p = .003$, $\eta^2_p = .29$) and HR increased ($F(1, 27) = 12.9$, $p = .001$, $\eta^2_p = .32$) significantly over 17 consecutive measuring points (i.e., 16 minutes incl. baseline T0; see Fig. 2 and 3). A significant difference of EDA ($F(2.33, 60.5) = 8.91$, $p = .0002$, $\eta^2_p = .26$) and HR ($F(5.26, 142) = 4.67$, $p = .0004$, η^2_p = .15) can be indicated with the highest increase in EDA after seven minutes into the experimental task (from $M = 0.44$, $SD = 1.94$ to $M = 1.53$, $SD = 2.16$; $t(26) = -4.08$, p = .0004; see Fig. 2, black dots) and the highest acceleration of the HR after six minutes into the experimental task (from $M = -0.001$, $SD = 0.04$ to $M = -0.016$, $SD = 0.036$; $t(27)$ = 2.71, $p = .012$; see Fig. 3, black triangles; HR acceleration means decreasing RR-intervals).

A distinctive feature can be observed after approximately six minutes (see Fig. 5 and 6): EDA and HR decline (i.e., RR curve rises) before increasing rapidly. At this time, the video ended and participants started reading the text, resulting in an attentional shift and a sudden increase in emotional activation (Lang, 2014). Afterward, the EDA and HR curves rose less sharply.

Fig. 5

EDA Changes With the Most Significant Increase From T7 to T8 (Black Dots)

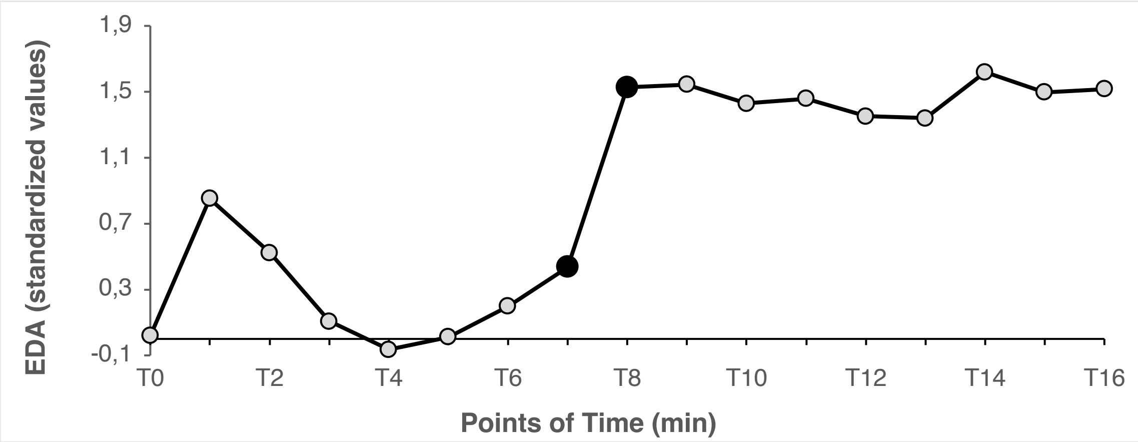
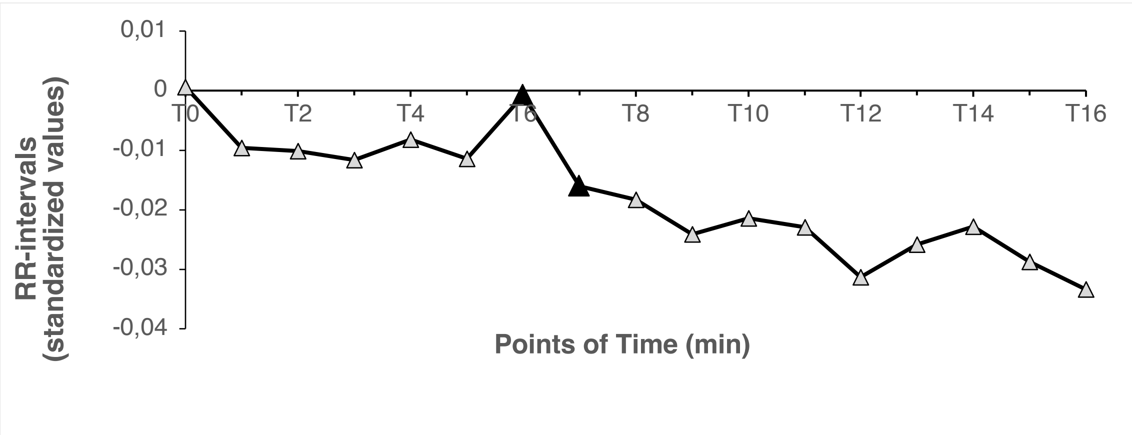


Fig. 6

HR Changes, Visualized in RR-intervals, With the Highest Decrease From T6 to T7 (Black Triangles). Decreasing RR-intervals Mean HR Acceleration



To examine whether the PNS or SNS controlled the HR, a spectral analysis for HRV was conducted using the *PhysioData Toolbox*. There, percentages were calculated to illustrate which nervous system was more active. The results showed that the low-frequency power (i.e., SNS) is 62.9 percent in charge of HR changes. At the same time, the high-frequency power (i.e., PNS) had only 29.8 percent control over HR. The remaining 7.23 percent corresponded to very low-frequency power and is associated with thermoregulation and, therefore, negligible.

Trend analyses showed a significant linear relation for EDA and learning performance (F(2, 24) = 4.10, p = .029, η^2_p = .26) supported the finding that higher learning scores go along with low EDA (see Fig. 7, on the right). For HR data, no statistically significant relation to learning performance can be reported (F(2, 25) = 1.05, p = .37). However, a visual inspection showed a linear trend between decreasing HR and increasing learning performance (see Fig. 7, on the left). Therefore, we conducted a One-Way ANOVA, resulting in significant difference between the groups of high, middle, and low learning performance for HR (F(2, 25) = 52.6, p < .001, η^2 = .81) and EDA (F(4.99, 59.9) = 2.30, p = .043, η^2 = .161; descriptive information in Table 4).

Fig. 7

Groups of High, Middle, and low Learning Performance to the Average of 17 Data Points of RR-intervals and EDA. A Decreasing RR-Interval Curve Means HR Acceleration

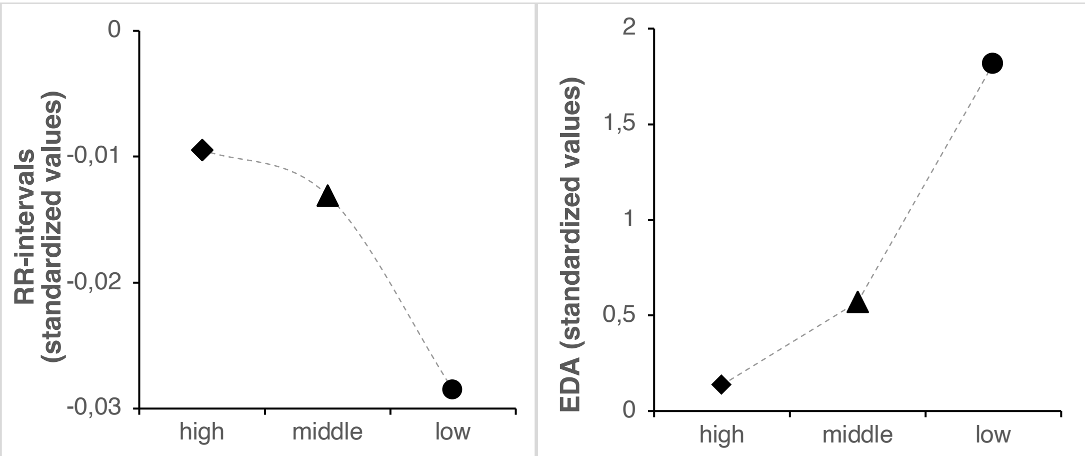


Table 4

Groups of High, Middle, and low Learning Performance for Electrodermal Activity (EDA) and Heart Rate (HR)

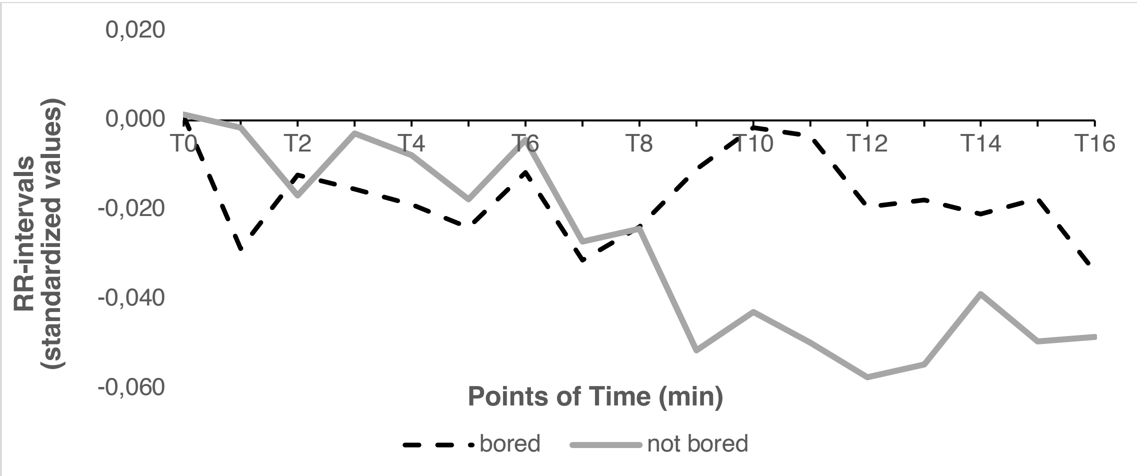
		N	*M*	*SD*
high	EDA	8	14.8	2.61
	HR	9	14.9	2.47
middle	EDA	9	10.2	0.67
	HR	9	10.3	0.71
low	EDA	10	5.60	2.22
	HR	10	5.60	2.22

Note. EDA = Electrodermal Activity, HR = Heart Rate.

As an explorative analysis, we compared participants which scored very high or very low on self-reported emotion questionnaires and analyzed whether specific emotions show a distinct psychophysiological pattern. Therefore, we aggregated the psychophysiological data of participants with differential values greater or less than zero (post-pre) individual items (sample size varies per item). Noticeable is the behavior in HR between bored (which scored one point higher in the posttest; $n = 5$) and not bored participants (which scored 2 points ($n = 2$) and one point ($n = 5$) lower in the posttest; see Fig. 8). Here, in four consecutive data points the HR was significantly higher for not bored ($n = 7$) than bored ($n = 5$) participants: after eight ($F(1, 10) = 5.65$, $p = .039$, $\eta^2 = .38$), nine ($F(1, 10) = 5.66$, $p = .039$, $\eta^2 = .36$), 10 ($F(1, 10) = 6.06$, $p = .034$, $\eta^2 = .38$) and 12 ($F(1, 10) = 5.16$, $p = .047$, $\eta^2 = .33$) minutes into the experimental task.

Fig. 8

HR comparison for bored vs. not bored participants



Note. From T8, the curves drift apart. Not bored participants show a significantly higher HR. Decreasing RR-intervals mean increasing HR.

Discussion and implications

In this work, we assessed if psychophysiological data can be used as an indicator for emotional states during learning with CBLEs and therefore predict learning performance (e.g., Pekrun et al., 2011, 2017; Pekrun & Stephens, 2012). Our exploratory research question targets the discourse of whether objective, real-time measures (i.e., psychophysiological data) reveal more information about the learning process than subjective post hoc self-reports. Compared to self-reports, which give the result of a learning session, psychophysiological data can measure what happens during the entire learning session and give real-time information about the learner's physiological behavior and emotional state. In our work, psychophysiological measurements were particularly fruitful given the progression of emotional states and task appraisal during learning and the shifting between the growth and well-being pathway. Different patterns were assessed by comparing groups that scored very high versus low on academic emotion scales. The characteristics, increasing EDA and HR, which interfere with learning, were detected. In addition, high EDA indicated low learning performance. Thus, psychophysiological measurements provide deeper insights into how and when academic emotions develop during learning than solely interpret self-reports.

Since our research question is relatively comprehensive, we defined precise hypotheses: Emotionally negative and activating learning material causes a decrease in HR and an increase in EDA, but differ depending on students' learning performance (high, middle, low).

Negative activating academic emotions cause HR deceleration over time (H1)

HR and negative activating emotions (frustration, distress, anxiety, and anger) increased after the learning phase, but this pattern is not aligned with our first hypothesis. However, our results indicate that the valence of deactivating academic emotions was expressed in HR because bored participants showed a lower HR (i.e., higher RR-intervals) than less bored participants (see Fig. 8). This leads to the assumption that HR can measure valence but not for highly activating emotions. Based on the research about to connection of HR and valence (see chapter 1.3.2.), HR can be a valid measure for valence. However, our learning environment's emotionally stimulating situation should be considered because the activation of the SNS could have superimposed the PNS and HR deceleration (Lang et al., 2009; see chapter 1.3.2.). This is in line with our finding that boredom expresses in decreasing HR (and increasing RR-intervals). Moreover, HRV analyses showed that the SNS is mainly in control over HR, concluding that the learning material was highly emotionally activating and therefore overcame the PNS (see chapter 1.3.2.). In summary, our first hypothesis cannot be supported, but the results indicate that changes in HR can reflect changing emotional states of learners.

A second possible explanation for the HR increase during the learning task besides high emotional activation of the learner can be the high cognitive load. Cranford and colleagues (2014) showed that tasks that cause a high cognitive load led to a higher increase in HR than tasks that elicit a small cognitive load. Also, Haapalainen and colleagues (2010) showed that ECG data was one of the most valuable indicators for cognitive load. Our results point in the same direction that HR displays rather cognitive load than the valence of academic emotions in a highly activating learning environment. Adding a control group with no emotionally activating stimuli would clarify this ambiguity. Moreover, qualitative data (open-ended questions or interviewing participants afterward) could provide a remedy.

Negative activating academic emotions cause increasing EDA over time (H2)

EDA data followed a significant linear trend corresponding to HR data, which is, considering the self-report results, in line with our second hypothesis. Prior research (e.g., Kreibig, 2010; Eteläpelto et al., 2018) showed that high EDA values are indicators for emotionally high activation, which corresponds with our findings. Consequently, EDA can be used as a reliable measure for emotional activation during learning. However, EDA cannot determine the valence of academic emotions. Since activating and deactivating academic emotions can benefit learning, a measure for the valence is necessary. Herewith, the importance of measuring the valence of academic emotions becomes apparent. We showed that HR could not perform this task, at least in the context of learning. Therefore, more research is needed to identify a reliable indicator of valence for academic emotions.

Depending on the learning performance, overall HR and EDA differ (H3)

Taking learning performance into account, a promising correlation can be found: With increasing EDA, the learning performance decreases. The activating learning material triggered negative academic emotions, which expressed in increasing EDA and led to poor knowledge increase. Though the posttest's learning score was significantly higher, the prior knowledge was relatively low due to the topic. So, it is not surprising that participants achieved a higher score in the posttest. Besides, the motivation of the learners could have been very high to prevent failure, resulting in high learning performance (see chapter 1.2.). This methodological issue should be considered for future research by choosing a more common topic. However, three significantly different groups for learning performance were identified. Therefore, EDA is a credible indicator of learning performance. For HR data, no clear statistical correlation was found. A trend can be detected when observing the results visually: with increasing HR, learning performance decreases. As a result, our third hypothesis can partly be supported.

Contrary to our expectations, the RR curve remains constantly below baseline level, triggered by activating, engrossing, and emotional learning material (Lang et al., 2009). This is in line with the findings that HR increases in highly emotional learning settings (Eteläpelto et al., 2018). Intense emotions like anxiety activate the SNS, resulting in faster HR and increasing EDA (Eteläpelto et al., 2018; Kreibig, 2010; Levenson et al., 2017). Based on these findings, the intensity of the experienced emotion could be the reason why we could not measure HR deceleration according to H3. We did not expect the overpowering emotional activation triggered by our learning environment. Our results point in the direction that in an emotionally high activating learning environment, HR is more sensitive for measuring cognitive load. Information input and attention usually go along with HR deceleration. When activating emotions, mental work, or concentration on inner thoughts are involved, the heart speeds up (Lang, 2014). This leads to the understanding that our setting provides an activating and emotional learning environment, which activates the SNS resulting in increasing EDA and HR. The activation of the SNS of our learning material overcomes the activation of the PNS, which slows the heart down (Lang et al., 2009).

An issue that remains to be discussed is the dramatic increase from T7 to T8 in EDA and from T6 to T7 in HR (see Fig. 5 and 6), which is a typical psychophysiological pattern for orienting responses. The reason behind an orienting response is the appearance of an unexpected stimulus (e.g., the sudden appearance of an error message on the screen or unexpected doorbell or call), which does not fit the current mental model (e.g., Bradley, 2009; Liebold et al., 2017; Potter & Bolls, 2012). This unexpected stimulus was the transition from the video to the text in our study. After the video stopped, the screen turned white before the text appeared. Moreover, the task shifted from watching the video passively to interacting with the input device (e.g., zooming the text in or out) and reading actively. Also, the participants' posture changed, from leaning back to sitting upright and closer to the screen. The EDA increased later than the HR because the electrodermal system is slower than the cardiovascular system (Berntson et al., 2017; Dawson et al., 2017). Since an orienting response refers to a short period and abates after a few seconds (Bradley, 2009), it has no further impact on our investigation.

Our overarching aim is to promote learning with CBLEs and find an implicit and real-time measurement for learning performance. Our research contributes to this issue by investigating how academic emotions manifest in psychophysiological data and validating physiological variables (e.g., EDA or HR) as a measurement of learning performance. The results are two options supporting the learner: as soon as a destructive academic emotion appears (e.g., frustration, boredom, anger (see Table 1), indicated by fluctuating and high EDA and HR), the learner receives support to solve the problem, prevent a switch to the well-being pathway (see chapter 1.1.), and lead the learner to learning success. The second assistance is identifying positive academic emotions (indicated by a steady EDA and HR), maintaining them, and keeping the learner on the growth pathway (see chapter 1.1.). Consequently, the learner's individual needs can be considered without getting out of the flow (Arguel et al., 2017).

Our findings and prior research on the significance and performance of psychophysiological measures show that it is worth establishing these measurements in CBLEs. An early approach to assessing emotions via an input device was "The Emotion Mouse" (Ark et al., 1999), which has not gained further acceptance because of the intrusive hardware. Since the technical state of the art nowadays is more sophisticated (e.g., smartwatches, fitness, or activity trackers), it is simple and unobtrusive to include these devices in CBLEs.

Limitations

Regarding our sample, gender differences can be noticed (see chapter 2.1.), which should be considered regarding the interpretation of the results of the self-reports. However, in a meta-analysis on emotions in technology-based learning environments, Loderer and colleagues (2020) could only find a weak relation between gender and academic emotions. Moreover, Frenzel, Pekrun, and Goetz (2007) showed that gender had no direct effect on academic emotions. Consequently, despite the gender differences, our sample can be considered reliable. Due to drop-outs, no noteworthy gender differences resulted in physiological data.

Although the task triggered negative and learning-inhibiting academic emotions, learners showed a significant knowledge increase. However, attention should be paid to the low prior knowledge, ensuring a higher posttest score. In future studies, the content of the learning material should be considered to clarify further connections of EDA and HR with learning success. Identifying relevant areas turned out to be difficult as we cannot be sure that all learners read the same text passage simultaneously. Previously defined areas or controlled reading speed could counteract this issue. The resulting comparable sections are more manageable in data processing and interpretation than looking at the learning session overall.

More research is needed to determine psychophysiological patterns for successful learning processes besides emotional and activating learning environments. Furthermore, the technical implementation of psychophysiological measurements and processing in digital environments is uncertain.

Conclusions

CBLEs gained importance, especially during the COVID-19 pandemic. However, learners' emotional states can hardly be identified by teachers in CBLEs. By making academic emotions measurable, learning progress can be better understood. The added value of this work is to comprehend the physiological appearance and impact of academic emotions on learning behavior and ultimately derive design approaches for CBLEs. In addition, this work aims further to validate psychophysiological measurements in the context of CBLEs, as this is relatively unattended (see Loderer et al., 2020).

Our findings show that psychophysiological measurements represent changes in academic emotions. Especially the distinction between the physiological behavior of bored and not bored participants can show shifting from Boekaerts' (2011) growth to well-being pathway. Bored participants chose the emotionally deactivating well-being pathway, especially with the increasing duration indicated by lower HR.

In conclusion, we found the physiological pattern of increasing HR and EDA, which indicates negative activating emotional states of learners in academic settings and EDA as sufficient indicator for learning performance. However, self-reports are essential at this stage of research to identify individual emotional states. Based on our research, it is possible to head in the direction of promoting learning using psychophysiological measurements.

More research is needed to combine knowledge about the physiological emergence of emotions, the connection to the physiological appearance of academic emotions, and learning processes. Currently, these are rather separate research areas but would enormously benefit from each other. Moreover, qualitative data (e.g., interviews, open-ended questionnaires, or think-aloud data) can be included to extend the findings and contribute to the multimodal data approach (see Järvelä et al., 2019).

References

Arguel, A., Lockyer, L., Lipp, O. V., Lodge, J. M., & Kennedy, G. (2017). Inside Out: Detecting Learners' Confusion to Improve Interactive Digital Learning Environments. *Journal of Educational Computing Research*, *55*(4), 526–551. https://doi.org/10.1177/0735633116674732

Baker, R. S. J. d., D’Mello, S. K., Rodrigo, Ma. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, *68*(4), 223–241. https://doi.org/10.1016/j.ijhcs.2009.12.003

Barrett, L. F., & Russell, J. A. (1999). The Structure of Current Affect: Controversies and Emerging Consensus. *Current Directions in Psychological Science*, *8*, 10-14. https://doi.org/10.1111/1467-8721.00003

Berntson, G. G., Quigley, K. S., Norman, G. J., & Lozano, D. L. (2017). Cardiovascular Psychophysiology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 183–216). Cambridge University Press. https://doi.org/10.1017/9781107415782.009

Boekaerts, M. (1999). Motivated learning: Studying student* situation transactional units. *European Journal of Psychology of Education*, *14*(1), 41. https://doi.org/10.1007/BF03173110

Boekaerts, M. (2011). Emotions, emotion regulation, and self-regulation of learning. In B. J. Zimmerman & D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 408–425). Routledge/Taylor & Francis Group.

Boucsein, W. (2012). *Electrodermal activity* (2nd ed). Springer.

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*(8), 1017–1034. https://doi.org/10.1111/j.1469-8986.2012.01384.x

Bradley, M. M. (2009). Natural selective attention: Orienting and emotion. *Psychophysiology*, *46*(1), 1–11. https://doi.org/10.1111/j.1469-8986.2008.00702.x

Bruhn, D., & Wollenteit, U. (2018). Konventionelle Schweinehaltung und Tierschutzgesetz. *Natur und Recht*, *40*(3), 160–169.

https://doi.org/10.1007/s10357-018-3310-9

Cacioppo, J. T., & Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *American Psychologist*, *45*(1), 16–28. https://doi.org/10.1037/0003-066X.45.1.16

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2017). Strong Inference in Psychophysiological Science. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 3–15). Cambridge University Press. https://doi.org/10.1017/9781107415782.001

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Lawrence Erlbaum Associates, Inc.

Cranford, K. N., Tiettmeyer, J. M., Chuprinko, B. C., Jordan, S., & Grove, N. P. (2014). Measuring Load on Working Memory: The Use of Heart Rate as a Means of Measuring Chemistry Students' Cognitive Load. *Journal of Chemical Education*, *91*(5), 641–647. https://doi.org/10.1021/ed400576n

Dawson, M. E., Schell, A. M., & Filion, D. L. (2017). The Electrodermal System. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 217–243). Cambridge University Press. https://doi.org/10.1017/9781107415782.010

D'Mello, S., & Graesser, A. (2014). Confusion and its dynamics during device comprehension with breakdown scenarios. *Acta Psychologica*, *151*, 106–116. https://doi.org/10.1016/j.actpsy.2014.06.005

D'Mello, S., Lehman, B., Pekrun, R., & Graesser, A. (2014). Confusion can be beneficial for learning. *Learning and Instruction*, *29*, 153–170. https://doi.org/10.1016/j.learninstruc.2012.05.003

Duffy, M. C., Lajoie, S. P., Pekrun, R., & Lachapelle, K. (2018). Emotions in medical education: Examining the validity of the Medical Emotion Scale (MES) across authentic medical learning environments. *Learning and Instruction*, 101150. https://doi.org/10.1016/j.learninstruc.2018.07.001

Eteläpelto, A., Kykyri, V.-L., Penttonen, M., Hökkä, P., Paloniemi, S., Vähäsantanen, K., Eteläpelto, T., & Lappalainen, V. (2018). A multi-componential methodology for exploring emotions in learning. *Frontline Learning Research*, 6–36. https://doi.org/10.14786/flr.v6i3.379

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior*

Research Methods, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics —A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education*, *22*(4), 497–514. https://doi.org/10.1007/BF03173468

Goetz, T., & Hall, N. C. (2013). Academic boredom. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 311–330). Routledge/Taylor & Francis Group.

Greenwald, M. K., Cook, E. W., & Lang, P. J. (1989). Affective judgment and psychophysiological response: Dimensional covariation in the evaluation of pictorial stimuli. *Journal of Psychophysiology, 3*(1), 51–64.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125. https://doi.org/10.1007/s40881-015-0004-4

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-physiological measures for assessing cognitive load. *Proceedings of the 12th ACM International Conference on Ubiquitous Computing*, 301–310. https://doi.org/10.1145/1864349.1864395

IBM Corp. (2020). *IBM SPSS Statistics for Windows* (26) [Computer Software].

Ijsselsteijn, de Ridder, H., Freeman, J., & Avons, S. E. (2000). Presence: Concept, determinants and measurement. *Human Vision and Electronic Imaging*, *3959*, 520–529. https://doi.org/10.1117/12.387188

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2019). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*, 101203. https://doi.org/10.1016/j.learninstruc.2019.04.004

Järvenoja, H., Järvelä, S., & Malmberg, J. (2017). Supporting groups' emotion and motivation regulation during collaborative learning. *Learning and Instruction*, 101090. https://doi.org/10.1016/j.learninstruc.2017.11.004

JASP Team. (2020). *JASP* (0.12.2) [Computer software].

Kang, M. J., Hsu, M., Krajbich, I. M., Loewenstein, G. F., McClure, S. M., Wang, J. T., & Camerer, C. F. (2008). The Wick in the Candle of Learning: Epistemic Curiosity Activates Reward Circuitry and Enhances Memory. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1308286

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology, 84*(3), 394–421. https://doi.org/10.1016/j.biopsycho.2010.03.010

Krohne, H. W., Egloff, B., Kohlmann, C.-W., & Tausch, A. (1996). Untersuchungen mit einer deutschen Version der „Positive and Negative Affect Schedule“ (PANAS). *Diagnostica 42*(2), 139–156.

Laarni, J., Ravaja, N., Saari, T., Böcking, S., Hartmann, T., & Schramm, H. (2015). Ways to measure spatial presence: Review and future directions. In Lombard M., Biocca F., Freeman J., IJsselsteijn W., Schaevitz R. (Eds.) *Immersed in Media*. Springer. https://doi.org/10.1007/978-3-319-10190-3_8

Lang, A. (1994). *What can the heart tell us about thinking?* In A. Lang (Ed.), *LEA's communication series. Measuring psychological responses to media messages* (p. 99–111). Lawrence Erlbaum Associates, Inc.

Lang, A. (2014). *Measuring Psychological Responses To Media Messages*. Routledge.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). International affective picture system (IAPS): Technical manual and affective ratings. *NIMH Center for the Study of Emotion and Attention*, *1*(39-58), 3.

Lang, P. J., Greenwald, M. K., Bradley, M. M., & Hamm, A. O. (1993). Looking at pictures: Affective, facial, visceral, and behavioral reactions. *Psychophysiology, 30*(3), 261–273. https://psycnet.apa.org/doi/10.1111/j.1469-8986.1993.tb03352.x

Lang, A., Potter, R., & Bolls, P. (2009). Where psychophysiology meets the media: Taking the effects out of mass media research. In J. Bryant & M. B. Oliver (Eds.), *Media Effects: Advances in theory and research* (3rd ed., pp. 185–206). Routledge.

Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In M. S. Clark (Ed.), *Emotion* (pp. 25–59). Sage Publications, Inc.

Leppert, K., Koch, B., Brähler, E., & Strauss, B. (2008). Die Resilienzskala (RS) – Überprüfung der Langfrom RS-25 und einer Kurzform RS-13. *Klinische Diagnostik Und Evaluation*, *1*, 226–243.

Levenson, R. W., Lwi, S. J., Brown, C. L., Ford, B. Q., Otero, M. C., & Verstaen, A. (2017). Emotion. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of Psychophysiology* (4th ed., pp. 444–464). Cambridge University Press. https://doi.org/10.1017/9781107415782.020

Loderer, K., Pekrun, R., & Lester, J. C. (2020). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction*, *70*, 101162. https://doi.org/10.1016/j.learninstruc.2018.08.002

Magno, C. (2011). Validating the Academic Self-regulated Learning Scale with the Motivated Strategies for Learning Questionnaire (MSLQ) and Learning and Study Strategies Inventory (LASSI). *The International Journal of Educational and Psychological Assessment*, *7*.

Palomba, D., Angrilli, A., & Mini, A. (1997). Visual evoked potentials, heart rate responses and memory to emotional pictorial stimuli. *International Journal of Psychophysiology*, *27*(1), 55–67. https://doi.org/10.1016/s0167-8760(97)00751-4

Panadero, E. (2017). A Review of Self-regulated Learning: Six Models and Four Directions for Research. *Frontiers in Psychology*, *8*, Article 422. https://doi.org/10.3389/fpsyg.2017.00422

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48. https://doi.org/10.1016/j.cedpsych.2010.10.002

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist*, *37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Pekrun, R., & Stephens, E. J. (2012). Academic emotions. In K. R. Harris, S. Graham, T. Urdan, S. Graham, J. M. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors.* (pp. 3–31). American Psychological Association. https://doi.org/10.1037/13274-001

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: The Epistemically-Related Emotion Scales. *Cognition and*

Emotion, *31*(6), 1268–1276. https://doi.org/10.1080/02699931.2016.1204989

Pinel, J. P. J., & Pauli, P. (2012). *Biopsychologie* (8., aktualisierte Auflage). Pearson, Higher Education.

Potter, R. F., & Bolls, P. (2012). *Psychophysiological measurement and meaning: Cognitive and emotional processing of media*. Routledge.

Preckel, F., Zeidner, M., Goetz, T., & Schleyer, E. (2008). Female "big fish" swimming against the tide: The "big-fish-little-pond effect" and gender ratio in special gifted classes. *Contemporary Educational Psychology, 33*(1), 78–96. https://psycnet.apa.org/doi/10.1016/j.cedpsych.2006.08.001

Shaffer, F., & Ginsberg, J. P. (2017). An Overview of Heart Rate Variability Metrics and Norms. *Frontiers in Public Health*, *5*. https://doi.org/10.3389/fpubh.2017.00258

Sjak-Shie, E. E. (2019). *PhysioData Toolbox* (0.5) [Computer software]. https://PhysioDataToolbox.leidenuniv.nl

Slater, M. (2002). Presence and The Sixth Sense. *Presence: Teleoperators and Virtual Environments*, *11*(4), 435–439. https://doi.org/10.1162/105474602760204327

Titz, W. (2001). *Emotionen von Studierenden in Lernsituationen: Explorative Analysen und Entwicklung von Selbstberichtskalen*. Waxmann.

Vermeer, H. J., Boekaerts, M., & Seegers, G. (2000). Motivational and gender differences: Sixth-grade students' mathematical problem-solving behavior. *Journal of Educational Psychology*, *92*(2), 308-315. https://doi.apa.org/doi/10.1037/0022-0663.92.2.308

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). Academic Press. https://doi.org/10.1016/B978-012109890-2/50045-7

Appendix B – Manuscript B (Study 2)

Investigating learning processes through analysis of navigation behavior using log files

Abstract

The empirical study investigates what log files and process mining can contribute to promoting successful learning. We want to show how monitoring and evaluation of learning processes can be implemented in the educational life by analyzing log files and navigation behavior. Thus, we questioned to what extent log file analyses and process mining can predict learning outcomes. This work aims to provide support for learners and instructors regarding efficient learning with computer-based learning environments (CBLEs). We evaluated log file and questionnaire data from students (N = 58) who used a CBLE for two weeks. Results show a significant learning increase after studying with the CBLE with a very high effect size (p <.001, g = 1.71). A cluster analysis revealed two groups with significantly different learning outcomes accompanied by different navigation patterns. The time spent on learning-relevant pages and the interactivity with a CBLE are meaningful indicators for Recall and Transfer performance. Our results show that navigation behaviors indicate both beneficial and detrimental learning processes. Moreover, we could demonstrate that navigation behaviors impact the learning outcome. We present an easy-to-use approach for learners as well as instructors to promote successful learning by tracking the duration spent in a CBLE and the interactivity.

Keywords: learning processes, navigation behavior, log files, computer-based learning environments, process mining

Theoretical background

The COVID-19 pandemic has demonstrated clearly that constant feedback and evaluation of the learning process, as well as progress in computer-based learning environments (CBLEs), pose challenges to teachers and learners alike (Grewenig et al., 2021). It has been extensively researched that monitoring and regulating the learning process and the progress of learners are essential to successful learning outcomes (Azevedo & Gašević, 2019; Hattie, 2017; McLaughlin & Yan, 2017; Schneider et al., 2021). Thus, the learning process needs to be tracked and evaluated before it can be customized to individual learners, especially whenever CBLEs are used (Arguel et al., 2017; Paans et al., 2020).

Real-time measures, such as log files (Reimann et al., 2014), physiological data (Malmberg et al., 2019), think-aloud protocols (Lim et al., 2021), and eye-tracking (Fan et al., 2022), have already proven beneficial as a way of analyzing learning processes; and they amount to a promising approach to providing instruction on a much-needed individual basis (Dindar et al., 2019; Goldman, 2009; Malmberg et al., 2019; Winne & Perry, 2000). However, in practice, these approaches have still not been implemented fully at educational institutions (Schneider et al., 2021).

The present work aims to show the extent to which log files can measure learning processes, as well as the impact that navigation behavior can have on learning. We present an easy-to-use method, which can be implemented in daily interactions with CBLEs.

Literature review

Collecting log files and exploring navigation behavior is a simple-to-implement, efficient method for tracing a learner's activity and interactions in CBLEs (Arguel et al., 2017; Cerezo et al., 2020; Huang & Lajoie, 2021; Matcha et al., 2019).

Monitoring the learners' interactions with a CBLE can provide insights into patterns of navigation behavior, which can influence feedback and teaching methods (Arguel et al., 2017; Azevedo & Gašević, 2019; Paans et al., 2020). This information can be gathered through log files and process mining, which are methods explored in the present work.

Measuring learning processes using log files

Log files record every interaction in a CBLE with a timestamp or the time spent on a particular page. Therefore, it is possible to analyze, for example, if the learner has carried out a specific task or read a learning-relevant text. In addition, the timestamp allows detection of how long the learners took for these activities. From this, it can be concluded that log files allow insights into individual learning and navigation behaviors (Azevedo et al., 2013; Malmberg et al., 2010). For example, systematic navigation behavior (frequent visits to learning-relevant pages) is positively correlated with increased knowledge (Bannert, 2006; Bannert et al., 2015). Lim and colleagues (2021) have shown that successful students re-read a learning-relevant text significantly more frequently than less successful students. Hence, navigation behavior is an indicator of learning outcomes.

Moreover, research has shown that monitoring of the learning process and Transfer performance correlate positively even after three weeks (Bannert et al., 2015; Sonnenberg & Bannert, 2015), which indicates that categories of learning (i.e., Recall, Comprehension, Transfer) correspond to navigation behavior. Furthermore, these categories are a crucial guideline for instructors when planning instructions and formulating learning goals (Anderson & Krathwohl, 2001; Krathwohl, 2002). For example, instructors can use flashcards to measure whether learners recall specific information. Writing a summary or blog journaling can be applied to measure if learners understand a concept. Additionally, instructors can instruct the learners to discuss an application example in chatrooms to determine if they can transfer their knowledge to new subjects (Churches, 2008). In order to explore if navigation behaviors reflect a specific category of the learning process, we developed a knowledge test, which measures each category but can also be summed up as a total learning score (see section 2.3.). Since these categories are structured from simple to complex (Bloom et al., 1956), we use the term *difficulty levels*, with the category *Recall* as the easiest, *Comprehension* as the intermediate, and *Transfer* as the hardest difficulty level.

Although log files are not as fine-grained as think-aloud data, they are an objective, automated measure. Thus, the learning process can be monitored without disturbing the learner (Hadwin et al., 2007; Winne, 2013). To analyze the sequence of events tracked in log files, we conducted a process mining model.

Describing the learning process using process mining

A popular analytical method for detecting patterns in navigation behavior is process mining (e.g., Lim et al., 2021; Sonnenberg & Bannert, 2016, 2019). Here, a process model is generated from log file data, which visualizes the interactions within the CBLE, based on specific events, revealing possible patterns of navigation behavior (Bannert et al., 2014). Based on these patterns, different groups of learners or learning strategies can be identified (e.g., Bannert et al., 2014; Huang & Lajoie, 2021; Matcha et al., 2019); thereby leading to the identification of either beneficial or detrimental learning behavior. This identification would make it possible to give individual, adequate feedback on the spot.

Consequently, we use navigation behavior to identify learning processes and also to ascertain whether it is possible to define groups of learners. We use log files to attain in-depth insights into learning behavior in combination with pre-post data (i.e., knowledge tests before and after learning). In view of the fact that we sought to present implementation in an everyday educational setting, we have evaluated data from a real seminar course.

Methods

The general question with which this study is concerned is which in-depth insights log files provide regarding learning behavior in a real seminar course, in conjunction with pre-post questionnaire data. Hence, we devised the following research questions and hypotheses:

- To what extent can navigation behavior predict learning outcomes? (RQ1)
    - Navigation behavior affects learning outcomes. (H1)
    - Navigation behavior reflects the difficulty level of the learning process. (H2)
- To what extent do learners differ, based on navigation and learning behaviors? (RQ2)
    - Learners with high learning outcomes display different patterns of navigation than learners with low learning outcomes. (H3)

To answer the research questions formulated, we monitored and evaluated a unit (14 days long) of long-term use over ten weeks of a CBLE in a real seminar at the University of Saarland in Germany. This seminar lasted from May until July 2020 and consisted of four online meetings and four learning units. The online meetings took place every two or three weeks, followed by intermediate self-studying phases. Each learning unit included a self-studying phase and a subsequent online meeting. We evaluated one learning unit consisting of a 14-day self-studying phase and one online meeting. The tasks for the self-studying phases were to acquire the respective content, read an additional scientific article, and write a summary about it; which then had to be uploaded for monitoring purposes. The CBLE represented the learning material during self-study phases, while the online meetings served as an opportunity to pursue dialogue with the teacher and to clarify any potential questions. There were no instructions for the students on logging in and out of the CBLE. Moreover, the study time was not prescribed by the teacher. Hence, the students had the freedom and flexibility to learn according to their preferences (i.e., when and for how long they wanted to learn).

The teacher introduced the CBLE, the procedure, and the seminar topics in the first online session. Subsequently, students filled in a pre-test to measure their prior knowledge regarding the topics addressed; this included a declaration of consent and information on the processing and retention of personal data. After the meeting, the first phase of self-study started. These online meetings and the self-study sequences were repeated four times in relation to four topics. In each online meeting, students completed the test of knowledge concerning the previously learned topic. For further data analyses, we focused on one particular topic, which showed the most significant increase in learning and a sufficient sample size. Moreover, it proved possible to reduce the inherent complexity of the process mining model used.

Participants

The participants consisted of 62 teacher-training students at the University of Saarland with a mean term time of 5.52 semesters ($SD = 1.83$) and with 41 females, 19 males, and one transgender individual. The mean age of the students was 22.18 years ($SD = 2.51$). Because four students did not complete the knowledge test after the self-study phase, only 58 participants could be included in the analysis of learning performances.

Learning environment

As the CBLE, the teacher used the *Toolbox TeacherEducation* (TTE), a German openly available, multimedia, and interactive learning platform for teacher-training students (Lewalter et al., 2018a). This contains scientific summaries of miscellaneous topics in teacher training (e.g., psychological - feedback; didactical - problem-solving, or subject-specific - Pythagorean theorem), video tutorials, staged videos about different teaching units, and tasks or questionnaires. The TTE has been used in real seminar courses since 2018 and is evaluated constantly to ensure that its use and content contribute to successful learning (Lewalter et al., 2018b, 2020, 2022; Titze et al., 2021).

Measures

The questionnaire (pre- and post-test) consists of 12 content-related multiple-choice items categorized in three difficulty levels based on *Bloom's Taxonomy* (1956): Recall, Comprehension, and Transfer (see section 1.1), and the hierarchically ordered *Thinking Skills* (Anderson & Krathwohl, 2001; Churches, 2008). The *Recall* level stands for remembering or recognizing facts. The *Comprehension* level refers to understanding and paraphrasing an issue. The *Transfer* level relates to designing and planning a new structure (Churches, 2008). Each difficulty level was measured in terms of four items (examples see Table 1). The total score is 58 points (Recall 20 points, Comprehension 20 points, Transfer 18 points). Using this classification, we were able to measure and distinguish between different skills.

The students were instructed that one or, indeed, none of the items might potentially be correct. In order to reduce the guess probability, each item offers the optional response "*I don't know*". The test was designed and validated, based on previous evaluations. It features an average Cronbach's alpha of .49, which is adequate, given that the questionnaires are designed explicitly for the TTE (see Schmitt, 1996; Taber, 2018). Since the TTE deals extensively with certain areas, we did not expect a high level of reliability overall (see Berger & Hänze, 2015).

Table 1
Example Items for Each Difficulty Level

Difficulty Level	Item Example
1: Recall	*"What are "open teaching" methods?"*
2: Comprehension	*"The teacher prepares a lesson in which learners have a high degree of choice regarding methods, media, and social format. Which approach does the teacher adopt?"*
3: Transfer	*"A teacher wants to use "tutorial learning" as a form of adaptive teaching. What should the teacher pay attention to?"*

The navigation behavior of the students was logged using the plugin *matomo* (https://matomo.org). Here, visits, visit durations, user ID, actions, page URLs, actions per visit, downloads, searches, transitions, and more can be tracked. We used duration, actions, page URLs, and user IDs for further data processing.

Data analysis procedure

The log files generated from *matomo* included user ID, duration, type of activity (e.g., click or download), page URL, page title, and a timestamp. The remaining variables were not used for further data analyses. To obtain a clearer picture, we labeled every page of the TTE with a simple acronym, which indicated the topic and the page order (e.g., topic 1, page 4 = t1_p04). Moreover, we categorized the pages into *learning-relevant* (pages with learning-related content, depending on the topic), *orienting* (i.e., dashboard, profile, settings, home), *learning-irrelevant* (text, videos, or tasks about topics unrelated to learning) and *videos* (pages that show videos exclusively).

We developed a Python script for automated data analysis of the following steps. We aggregated the time spent on the categorized pages and this resulted in the variables *learning-relevant*, *learning-irrelevant, orienting*, and *videos* for each visit. Next, we summarized each log file per student. This resulted in a data set, which included all of the students and the respective duration, learning-relevant, orienting, learning-irrelevant times, and videos. We processed the log files for process mining techniques, using the pm4py Python package. The resulting output included a visit ID (user ID and visit count), activity (page acronym), and a timestamp (duration spent on the page). For process mining, we used the software application *Disco* from Fluxicon (https://fluxicon.com/disco/).

Results

The following results are clustered, based on our data analysis procedure. Initially, we present descriptive data and a declaration of essential variables (see Tables 2 and 3). Additionally, we ran a paired samples t-Test (prior knowledge - learning outcome) to examine whether knowledge increased significantly after studying with the TTE. Afterwards, we present results from bivariate correlations to confirm H1, H2, and to answer RQ1 (see Table 4). Next, we show a hierarchical cluster analysis, including a One-Way ANOVA, in order to address H3 and RQ2 (see Table 5 and Fig. 1). Based on the resulting clusters, we present our process mining approach, to give an exhaustive answer to RQ1 and RQ2 (see Tables 6, 7, and Fig. 2).

Table 2

Important Variable Names and Their Declaration

	Variable Name	Declaration
Prior knowledge	pre-Difficulty Level 1	total pre-test score for questions for Recall (see Table 1)
	pre-Difficulty Level 2	total pre-test score for questions for Comprehension (see Table 1)
	pre-Difficulty Level 3	total pre-test score for questions for Transfer (see Table 1)
	Pre-test score	total pre-test score for all questions
Learning Outcome	Difficulty Level 1	total post-test score for questions for Recall (see Table 1)
	Difficulty Level 2	total post-test score for questions for Comprehension (see Table 1)
	Difficulty Level 3	total post-test score for questions for Transfer (see Table 1)
	Post-test score	total post-test score for all questions
Navigation Behavior (NB)	duration (s)	time spent in the learning environment
	actions	number of mouse clicks within the learning environment
	learning-relevant (s)	time spent on pages where theoretical basics are defined (e.g., models, state of research, concepts)
	orienting (s)	time spent on pages that serve the orientation in the learning environment (e.g., overview, profile)
	learning-irrelevant (s)	difference value of duration - time learning-relevant
	videos (s)*	time spent on pages where solely videos are displayed

Note. Time and duration variables are measured in seconds.
*Only relevant for cluster analyses.

At first, we identified any outliers, using z-scores, resulting in different sample sizes (see Table 3). The mean score for overall prior knowledge (i.e., the pre-test score) was 32.98 and, therefore, above half of the maximum achievable score of 58 (see Table 3). Equally, the mean scores for Difficulty Levels 1 (M = 12.05) and 2 (M = 11.57) were above half the maximum score of 20. The mean score for Difficulty Level 3 was 8.74 and almost half the maximum achievable score of 18 (see Table 3). The most significant increase of 5.16 points was measured for Difficulty Level 3 and the smallest increase of 1.38 points for Difficulty Level 2. The knowledge gain in total was 8.67 points.

To analyze if the knowledge scores increase from prior knowledge (i.e., pre-test score) to the learning outcome (i.e., the post-test score) was significant, we carried out a paired samples t-Test. The results showed that the scores for all three Difficulty Levels (Level 1: $t(41) = 6.26$, $p < .001$, $g = 1.12$; Level 2: $t(41) = 2.67$, $p = .005$, $g = 0.55$; Level 3: $t(41) = 10.43$, $p < .001$, $g = 2.05$) and the scores in total ($t(40) = 8.06$, $p < .001$, $g = 1.71$) increased significantly with medium to very high effect sizes.

Table 3

Descriptive Data for Learning Outcome and Navigation Behavior

	Variable	n^*	Min	Max	*M*	*SD*
Prior knowledge	pre-Difficulty Level 1	42	3	18	12.05	2.95
	pre-Difficulty Level 2	42	4	16	11.57	3.05
	pre-Difficulty Level 3	43	3	13	8.74	2.85
	Pre-test score	41	13	42	32.98	7.65
Learning Outcome	Difficulty Level 1	43	10	20	14.95	2.35
	Difficulty Level 2	43	7	19	12.95	2.65
	Difficulty Level 3	42	8	17	13.90	2.15
	Post-test score	43	28	56	41.65	5.61
Navigation Behavior (NB)[a]	duration (s)	43	42	11289	4298	3155
	actions	43	0	53	22.74	13.23
	orienting (s)	44	32	3263	1009	891.9
	learning-relevant (s)	44	0	8821	2999	2460
	learning-irrelevant (s)	43	32	3453	1213	985.5

Note. Time and duration variables are measured in seconds.
[*]Outliers were identified and excluded for further data analyses, resulting in different sample sizes.
The maximum score for learning outcome is 58 in total.
The maximum score for Difficulty Levels 1 and 2 is 20 and for Difficulty Level 3, 18.
[a]The variable "video" is not included here because it was only relevant for cluster analyses.

Because our first hypothesis (see section 2) addresses the influence of the navigation behavior (i.e., duration, actions, orienting, learning-relevant, learning-irrelevant) and learning outcome (i.e., post-test score), we analyzed the relationships between these constructs (see Table 4). The variable "video" is not included here because it was only relevant for cluster analyses. The bivariate correlation analysis indicated that, except for the variables orienting and learning-irrelevant, navigation behavior has a positive linear relationship with the post-test score. Thus, navigation behavior affects the learning outcome (see Table 4; results can be seen in the second row). However, prior knowledge has no significant relationship with navigation behavior (see Table 4; results can be seen in the first row).

To address the more detailed H2, we examined the correlations described for each Difficulty Level separately. In this way, it was possible to map out the difficulty level of the learning process (see section 1.1). Results show that Difficulty Level 2 does not correlate with any variable relating to navigation behavior. However, the variables duration, actions, and learning-relevant have a positive linear relationship with Difficulty Levels 1 and 3. The variables orienting and learning-irrelevant do not correlate with learning outcome (see Table 4).

Table 4

Results for Correlation Analysis of Learning Outcome and Navigation Behavior

	duration[a]		actions[a]		orienting[b]		learning-relevant[b]		learning-irrelevant[b]	
	p	*r*	*p*	*r*	*p*	*r*	*p*	*r*	*p*	*r*
Pre-test Score	.378	-.143	.116	-.252	.789	-.043	.279	-.173	.875	.026
Post-test Score	.028	.338*	.028	.339*	.053	.297	.007	.404**	.472	.114
Difficulty Level 1	.021	.356*	.009	.398**	.232	.186	.006	.410**	.229	.190
Difficulty Level 2	.238	.186	.073	.280	.064	.285	.069	.280	.924	.015
Difficulty Level 3	.017	.371*	.049	.309*	.096	.260	.017	.366*	.567	.092

Note. Pearson correlation, 2-tailed, ** $p < .01$, * $p < .05$
[a]$n = 42$
[b]$n = 43$

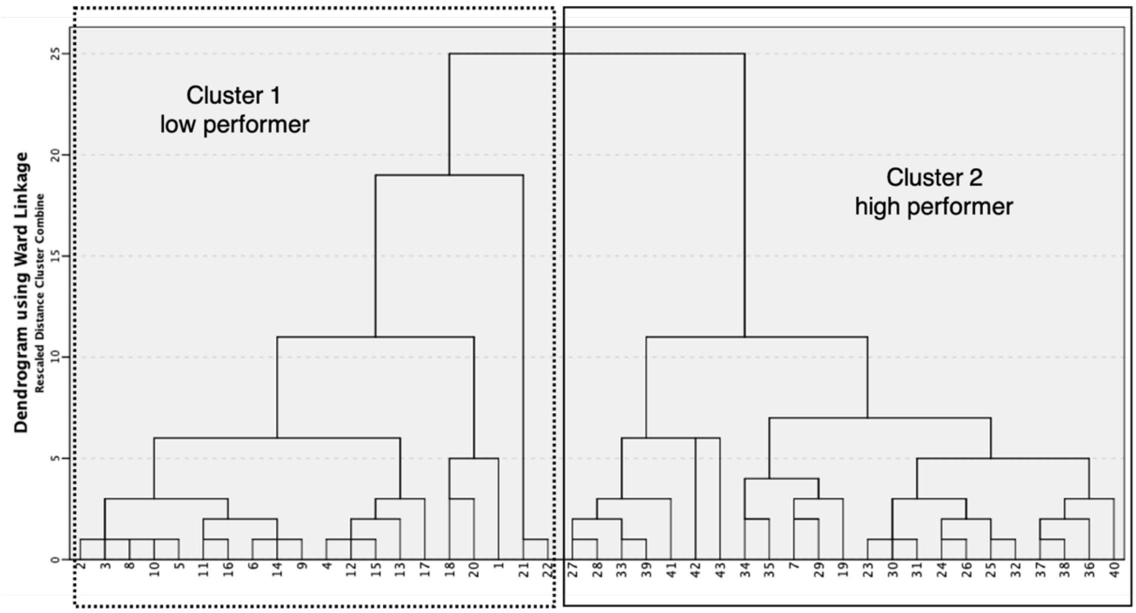
Since log files contain a large quantity of data and we consider our dataset promising, we wanted to zoom in and explore it in greater detail. The analyses performed above, which are rather conservative, were unable to uncover the dynamic, individual character of navigation behaviors. Therefore, we conducted a hierarchical cluster analysis, using the Ward Linkage; which generated highly homogeneous clusters (see also Huang & Lajoie, 2021; Paans et al., 2020; see Fig. 1).

In order to detect distinct groups of learners (see H3), we included all outliers in the data set. Only one participant had to be excluded due to missing post-test data, resulting in a sample of $N = 43$. The dendrogram from the hierarchical cluster analysis showed two meaningful clusters (see Fig. 1; cluster 1: $n = 20$, cluster 2: $n = 23$). The x-axis represents the anonymized number of each participant. The y-axis shows the distance between each cluster procedure (see Fig. 1). For the analysis, z-scores were used, though the dendrogram presents raw scores. We chose the Euclidean distance as the distance measure; and, due to the sample size, we predefined two clusters, in order to get a clear division. We used the navigation behavior (i.e., duration, actions, orienting, learning-relevant, learning-irrelevant, and videos) as cluster variables. Based on this procedure, we managed to identify two distinct groups (see Fig. 1). Since this division seemed sufficient and was of a similar sample size, we proceeded with the suggested clusters. Afterward, we conducted a variance analysis to examine how the groups differ based on their navigation behavior and learning outcomes and whether this difference is significant (see Table 5).

Both groups differ significantly in their navigation and learning behaviors (see Table 5). Noteworthy is the fact that the first cluster had consistently fewer values for all variables (Fig. 1, left side, and Table 5). However, because the prior knowledge in this case was not significantly different from the "better" group, we named the first cluster *low performers* and the group with higher values *high performers*.

Fig. 1

Dendrogram Visualizing two Meaningful Cluster



The low performers showed a significantly lower learning outcome on all three Difficulty Levels, especially on Level 1 and 3, with very high effect sizes (see Table 5). The group differences could also be measured in the navigation behavior. The high performers spent more than twice as much time in the TTE on learning-relevant, orienting, and video pages and performed almost twice as many actions as the low performers. However, the high performers also visited learning-irrelevant pages significantly longer than the low performers. Since the high performers showed a higher duration overall without significant differences for learning-irrelevant pages, this is not further questionable (see Table 5).

Table 5

Means, Standard Deviations, and One-Way Analysis of Variance in Low and High Performers

Variable	low performers[a]		high performers[b]		$F(1, 41)$	η^2
	M	*SD*	*M*	*SD*		
Pre-test score	31.1	9.16	32.5	6.20	0.341	.008
Difficulty Level 1	13.5	1.761	16.4	1.90	26.7^{***}	.394
Difficulty Level 2	11.7	2.03	14.0	2.69	10.2^{*}	.199
Difficulty Level 3	12.4	2.26	14.8	1.85	15.0^{***}	.268
Post-test score	37.6	4.50	45.2	3.74	37.2^{***}	.476
duration	1914	2197	5583	2835	22.0^{***}	.349
actions	15.7	12.6	31.3	12.1	17.2^{***}	.295
orienting	516	598	1175	898	7.75^{*}	.159
learning-relevant	1530	1745	4387	2220	21.5^{***}	.344
learning-irrelevant	384	667	1196	1697	4.03	.089
videos	137	429	974	1499	5.80^{*}	.124

Note. $^{*}p < .05$, $^{***}p < .001$

$^{a}n = 20$

$^{b}n = 23$

Besides detecting different learner groups through log file and cluster analyses, we were interested in examining the sequence and flows of the navigation behavior of each learner group. Therefore, we conducted process mining analyses, which are a fruitful approach to reveal such sequential flows. Here, we used the log file data to create a process mining model based on the significantly different clusters that resulted. We summarized homogeneous pages (e.g., pages with text or tasks) to generate a transparent process model (see Table 6).

Table 6

Variables for Process Modeling and Declaration

Variable Name	Declaration
text	pages with theoretical basics, concepts, and models
video	pages with videos exclusively
task	pages with tasks/questionnaires
literature	list of references
orienting	pages that serve the orientation in the learning environment

The results from the process analyses support the cluster analysis carried out and the emerging groups of low and high performers demonstrated (see Fig. 1 and Table 5). The high performers visit text, video, task, and literature more than twice as often as the low performers. Moreover, high performers visit orienting pages more often than the low performers, though the difference is not as significant as with the other categories (see Table 7 and Fig. 2).

Table 7

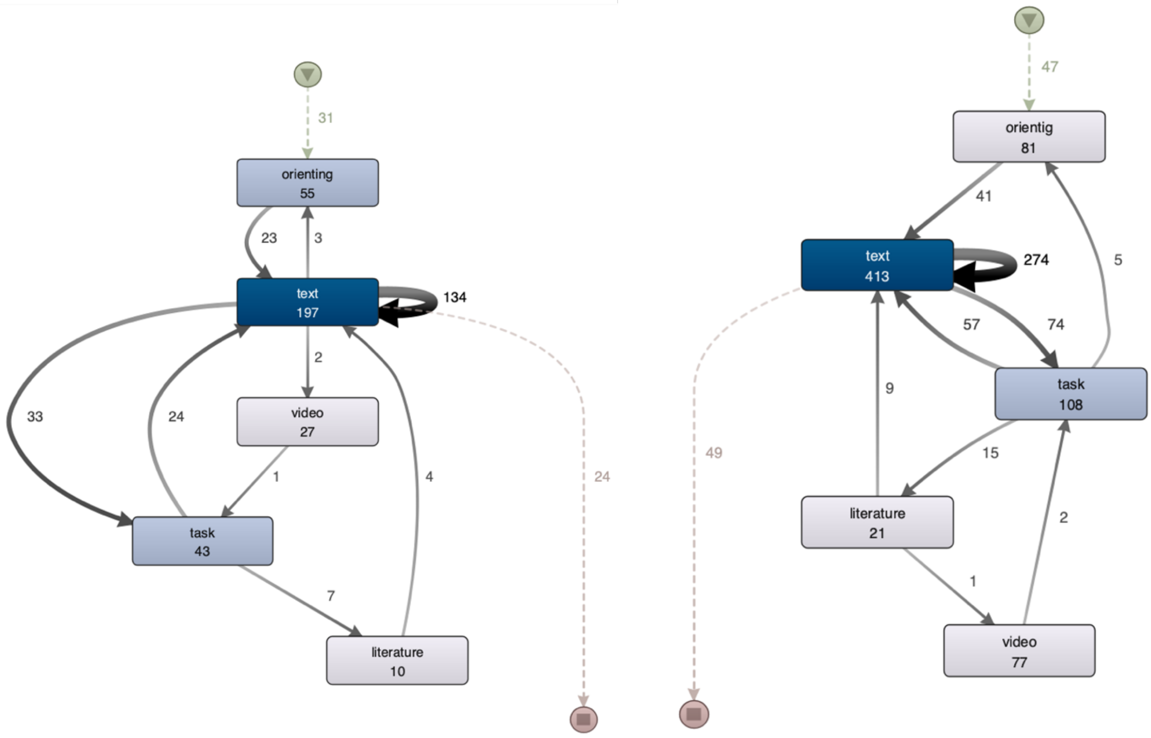
Absolute and Mean Activity Frequency for Each Category

Categories	Activity Frequency			
	Low Performers		High Performers	
	Absolute Value	*M*	Absolute Value	*M*
text	197	9.85	413	17.96
video	27	1.35	77	3.35
task	43	2.15	108	4.70
literature	10	0.50	21	0.913
orienting	55	2.75	81	3.52

After presenting the descriptive data, we exported the process model to illustrate low and high performers' tread routes (see Fig. 2). Both group models start with orienting pages and walk along to text pages. It is noticeable that there is a loop on the process model showing the high performers. They go directly back and forth from the text to task pages, probably in order to verify their knowledge; and then they return to learning-relevant content on the text pages. Low performers, on the other hand, go directly to video, task, and literature pages without any major loop pattern (see Fig. 2).

Fig. 2

Process Model for Low Performers (left) and High Performers (right)



Discussion and implications

In this study, we investigated the extent to which log files and navigation behavior can predict learning outcomes. Moreover, we sought to show that learners with high learning outcomes display different navigation behavior than learners with low learning outcomes (e.g., Bannert, 2006; Lim et al., 2021).

Our approach contributes to meeting the challenge of how instructors can monitor and evaluate the learning process and progress in CBLEs; and how they can give adequate feedback at the appropriate time, based on the learner's needs (Paans et al., 2020; Schneider et al., 2021). Thus, our approach and findings can also support instructors and designers of CBLEs. Instructors can observe how learners interact with the CBLE and what individual learning style they prefer for effective learning. Thus, instructors can mitigate the absence of face-to-face interaction, given that immediate feedback is more effective (Bloom, 1984). Based on instructors' observation of learners, beneficial learning methods can be introduced or promoted (Hillmayr et al., 2020; Van der Kleij et al., 2015). If a specific navigation pattern leads to low learning outcomes, the designer of the CBLE would be able to adapt the learning content, environment, or instructions in such a way as to ensure successful learning (Bousbia et al., 2010).

We hypothesized that navigation behavior affects the learning outcome. Therefore, we correlated the post-test score with navigation behavior. The results obtained show that the post-test score has a significant linear relationship with the variables - duration, actions, and learning-relevant (see Table 4). Thus, higher duration, especially on learning-relevant pages, as well as a greater number of actions, contribute to successful learning; which supports our first hypothesis. These results give rise to the conclusion that learners need to invest time on learning-relevant pages and engage actively with the CBLE to reach high-level learning outcomes; which is in line with the findings of Mayer (2014). The pre-test score does not correlate significantly with the navigation behavior. Hence, prior knowledge does not affect navigation behavior.

Our second hypothesis addressed the relation between the difficulty levels of learning (Recall, Comprehension, Transfer) and navigation behavior. Here, we wanted to analyze if navigation behavior can reflect Recall, Comprehension, or Transfer performance (e.g., high Transfer performance goes along with a different navigation pattern than high Recall performance). In fact, we were able to show a significant linear relationship between Difficulty Levels 1 (Recall) and 3 (Transfer) and navigation behavior (i.e., duration, actions, learning-relevant, see Tables 1 and 4); which means that the longer the students stayed in the TTE and, especially, on learning-relevant pages, the better Recall and Transfer performance were. The time spent on orienting pages shows, as expected, no significant relationship with learning but, at the same time, does not negatively affect learning (see Table 4). Based on these results, our second hypothesis can also be supported.

Regarding our first research question, we were able to show that the time spent on learning-relevant pages and the associated interactivity (i.e., number of actions) are important factors for high learning outcomes. Moreover, here, conclusions regarding the level of difficulty can be drawn: The more extended learners stay on learning-relevant pages and the more intense their interaction (measured by actions) with the TTE, the better the Recall and Transfer performance. The implications for instructors are that the durations and actions within a CBLE are meaningful for successful learning. By marking learning-relevant pages in a CBLE and tracking the intensity of interactions, as well as the time learners spent in the CBLE, instructors can ensure high learning outcomes. Since these factors can be measured and evaluated quite rapidly, implementation in everyday education is, indeed, feasible.

We hypothesized that learners with high learning outcomes show a different navigation pattern than learners with a low learning outcome. To validate this third hypothesis and based on Huang and Lajoie (2021) and Paans and colleagues (2020), we implemented both a cluster analysis and a process model (see Fig. 1 and 2). The two resulting groups (low performers and high performers) differ significantly regarding the learning outcome and navigation behavior, which supports our third hypothesis. However, both groups show similar prior knowledge. A significant difference in Recall, Comprehension, and Transfer performance in favor of the high performers can be measured (see Table 5). Additionally, the high performers showed significantly higher durations on learning-relevant, orienting, and video pages (more than twice as long). Moreover, the high performers interacted more actively with the TTE (more than twice as many actions). Interestingly, the high performers also stayed more than twice as long on learning-irrelevant pages. However, this result is not unusual, since the high performers have an overall higher duration. An explanation for the poor interaction of the low performers could be that, due to their high level of prior knowledge, they did not see the need to acquire the learning content.

Regarding our second research question, namely, the extent to which learners differ, as measured by navigation behavior and learning outcome, we included a process model designed to make possible navigation patterns visible. This model reveals that the high performers show higher activity frequencies for text, video, orienting, literature pages, and tasks implemented in the TTE (see Tables 6 and 7). Lim and colleagues (2021), as well as Bannert and colleagues (2014), showed that high-frequency activity leads to superior learning outcomes. More precisely, they identified specific self-regulated learning phases by categorizing the activities involved (e.g., orientation, planning, monitoring, search, evaluation; Bannert et al., 2014; Lim et al., 2021). Doing so makes the actions and interactions with the CBLE more specific regarding self-regulated learning.

The process model reveals that the high performers also have a different pattern of navigation behavior, which leads to a superior learning outcome. The high performers present a conspicuous looping pattern, including the text pages and the tasks, which gives rise to the conclusion that they read a text, test their knowledge by carrying out a task, and then return to studying.

As with the high performers, MacGregor (1999) found patterns in learners' navigation behavior by conducting a cluster analysis: The "sequential studiers" are distinctive in terms of methodical strategy and focus on reading. Lawless and Kulikowich (1996) found users by performing a cluster analysis, which spent little time in the CBLE and did not use many features or inspected pages ("apathetic hypertext users"). This pattern is similar to our finding regarding the low performers.

In conclusion, we contributed to the research field of log file analyses and process mining approaches by showing that log files are a robust tool suited to obtaining information about the learning process in a CBLE. Additionally, we demonstrated that analyzing navigation behavior is a promising approach when it comes to predicting learning outcomes. We were able to demonstrate navigation behavior patterns that indicate both, beneficial (high performers) and detrimental (low performers) learning. Our work counters the absence of monitoring learners' activity in a CBLE; and it does this by presenting a method that is easy to use, easy to evaluate and easy to integrate into daily educational routines. Thus, instructors can detect either beneficial or detrimental learning processes, as appropriate, and then provide adequate feedback.

Limitations

Regarding learning-related variables, it is striking that the time learners spend on learning-irrelevant pages does not correlate negatively with learning outcomes. The actual generation of this variable provides an explanation: Given that we evaluated just one section of the entire semester, we needed to infer what was "learning-irrelevant" from within this section. Because as soon as we define every page beside the topic we analyzed as learning-irrelevant, a fuzzy and disproportionally large number remains. Thus, the variable learning-irrelevant could be inconclusive.

Our results show that Difficulty Levels 1 (Recall) and 3 (Transfer) are meaningful variables. Yet, Difficulty Level 2 (Comprehension) does not correlate with navigation behavior and thus, seems of no significance. The reason for this could be the nature of our self-designed questionnaire, which probably did not define the Comprehension category clearly enough. However, despite this, Recall and Transfer performance, as well as the overall post-test score, are meaningful indicators of knowledge gain and are sufficient for our purposes.

We mentioned the connection between self-regulated learning phases and activities within a CBLE (see “Discussion and implications”). Including self-regulated learning, various measures would present our variable actions more precisely and would yield information about learners’ cognitive processes. However, self-regulated learning and its impact on navigation behavior and learning outcomes have already been researched sufficiently (e.g., Bannert et al., 2014, 2015; Fan et al., 2022; Lim et al., 2021; Matcha et al., 2019; Schoor & Bannert, 2012; Sonnenberg & Bannert, 2015).

Conclusion

Learning with CBLEs is indispensable and provides a host of benefits for learners. Learners can interact actively with learning content and study at their own pace and in their preferred learning style. Due to their static, unified setting, both traditional lectures and frontal teaching methods have recently been called into question. Studying with CBLEs counters these issues, because learners can acquire knowledge in line with their own needs (Goedhart et al., 2019; Mingorance Estrada et al., 2019). However, instructors must track the learning process and learners’ progress itself, in order to ensure that specific learning goals are met.

Besides questionnaires, navigation behavior is a highly useful measure when it comes to tracking the learning process and associated progress (e.g., Bousbia et al., 2010; Matcha et al., 2019; Paans et al., 2019).

Our results show that both, log files and navigation behavior can predict learning outcomes: The time spent in a CBLE and the intensity of interactions with the CBLE yield information about Recall and Transfer performance. Furthermore, learners with a high learning outcome navigate differently through the CBLE (see Fig. 2): The high performers interact with the CBLE in a more frequent and intense manner. Moreover, the pattern of navigation behavior varies, depending on the learning outcome: High performers show a specific linkage between text and tasks (see Fig. 2).

Based on our research, log files and navigation behavior are validated to predict learning outcomes and the difficulty level of learning outcomes (Recall and Transfer performance). We were able to show that navigation behavior significantly impacts learning outcomes and that learners with high learning outcomes display significantly different navigation behavior than learners with low learning outcomes (see also Bannert, 2006; Lim et al., 2021).

Our approach and results are promising, since we evaluated data from a real seminar course and successfully tested the feasibility of implementing the monitoring of learning processes in a CBLE.

Acknowledgments: The digital learning platform *Toolbox TeacherEducation* is funded by the Federal Ministry of Education and Research as part of the joint quality initiative for teacher training by the federal and state governments.

References

Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman, Inc.

Arguel, A., Lockyer, L., Lipp, O. V., Lodge, J. M., & Kennedy, G. (2017). Inside Out: Detecting Learners' Confusion to Improve Interactive Digital Learning Environments. *Journal of Educational Computing Research*, *55*(4), 526–551. https://doi.org/10.1177/0735633116674732

Azevedo, R., & Gašević, D. (2019). Analyzing Multimodal Multichannel Data about Self-Regulated Learning with Advanced Learning Technologies: Issues and Challenges. *Computers in Human Behavior*, *96*, 207–210. https://doi.org/10.1016/j.chb.2019.03.025

Azevedo, R., Harley, J., Trevors, G., Duffy, M., Feyzi-Behnagh, R., Bouchet, F., & Landis, R. (2013). Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In *International Handbook of Metacognition and Learning Technologies* (pp. 427–449). Springer.

Bannert, M. (2006). Effects of reflection prompts when learning with hypermedia. *Journal of Educational Computing Research*, *35*(4), 359–375.

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, *9*(2), 161–185. https://doi.org/10.1007/s11409-013-9107-6

Bannert, M., Sonnenberg, C., Mengelkamp, C., & Pieger, E. (2015). Short- and long-term effects of students' self-directed metacognitive prompts on navigation behavior and learning performance. *Computers in Human Behavior*, *52*, 293–306. https://doi.org/10.1016/j.chb.2015.05.038

Berger, R., & Hänze, M. (2015). Impact of Expert Teaching Quality on Novice Academic Performance in the Jigsaw Cooperative Learning Method. *International Journal of Science Education*, *37*(2), 294–320. https://doi.org/10.1080/09500693.2014.985757

Bloom, B. S., Englehart, M., Furst, E., Hill, W., & Krathwohl, D. (1956). Taxonomy of Educational Objectives: Handbook I Cognitive Domain. *New York, New York:*

David McKay Company, 144–145.

Bousbia, N., Rebaï, I., Labat, J.-M., & Balla, A. (2010). Learners' navigation behavior identification based on trace analysis. *User Modeling and User-Adapted Interaction*, *20*(5), 455–494. https://doi.org/10.1007/s11257-010-9081-5

Cerezo, R., Bogarín, A., Esteban, M., & Romero, C. (2020). Process mining for self-regulated learning assessment in e-learning. *Journal of Computing in Higher Education*, *32*(1), 74–88. https://doi.org/10.1007/s12528-019-09225-y

Churches, A. (2008). Bloom's taxonomy blooms digitally. *Tech & Learning*, *1*, 1–6.

Dindar, M., Malmberg, J., Järvelä, S., Haataja, E., & Kirschner, P. A. (2019). Matching self-reports with electrodermal activity data: Investigating temporal changes in self-regulated learning. *Education and Information Technologies*. https://doi.org/10.1007/s10639-019-10059-5

Disco (3.0.0) [Software]. (2021). Fluxicon BV. https://fluxicon.com/disco/

Fan, Y., Lim, L., van der Graaf, J., Kilgour, J., Raković, M., Moore, J., Molenaar, I., Bannert, M., & Gašević, D. (2022). Improving the measurement of self-regulated learning using multi-channel data. *Metacognition and Learning*. https://doi.org/10.1007/s11409-022-09304-z

Goedhart, N. S., Blignaut-van Westrhenen, N., Moser, C., & Zweekhorst, M. B. M. (2019). The flipped classroom: Supporting a diverse group of students in their learning. *Learning Environments Research*, *22*(2), 297–310. https://doi.org/10.1007/s10984-019-09281-2

Goldman, S. R. (2009). Explorations of relationships among learners, tasks, and learning. *Learning and Instruction*, *19*(5), 451–454. https://doi.org/10.1016/j.learninstruc.2009.02.006

Grewenig, E., Lergetporer, P., Werner, K., Woessmann, L., & Zierow, L. (2021). COVID-19 and educational inequality: How school closures affect low-and high-achieving students. *European Economic Review*, *140*, 103920.

Hadwin, A. F., Nesbit, J. C., Jamieson-Noel, D., Code, J., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, *2*(2–3), 107–124. https://doi.org/10.1007/s11409-007-9016-7

Hattie, J. (2017). *Backup of Hattie's ranking list of 256 influences and effect sizes related to student achievement*.

Huang, L., & Lajoie, S. P. (2021). Process analysis of teachers' self-regulated learning patterns in technological pedagogical content knowledge development.

Computers & Education, *166*, 104169. https://doi.org/10.1016/j.compedu.2021.104169

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. Theory Into Practice, 41(4), 212–218. https://doi.org/10.1207/s15430421tip4104_2

Lewalter, D., Schiffhauer, S., Richter-Gebert, J., Bannert, M., Engl, A.-T., Maahs, M., Reißner, M., Ungar, P., & von Wachter, J.-K. (2018a). *Toolbox Lehrerbildung*.

Lewalter, D., Schiffhauer, S., Richter-Gebert, J., Bannert, M., Engl, A.-T., Maahs, M., Reißner, M., Ungar, P., & von Wachter, J.-K. (2018b). Toolbox Lehrerbildung: Berufsfeldbezogene Vernetzung von Fach, Fachdidaktik und Bildungswissenschaft. *Kohärenz in Der Universitären Lehrerbildung*, 331–353.

Lewalter, D., Schneeweiss, A., Richter-Gebert, J., Huber, K., & Bannert, M. (2022). Mit Unterrichtsvideos praxisnah und disziplinverbindend lehren und lernen. *Lehren Und Forschen Mit Videos in Der Lehrkräftebildung*, 125.

Lewalter, D., Titze, S., Bannert, M., & Richter-Gebert, J. (2020). Lehrer*innenbildung digital und disziplinverbindend. Die Toolbox Lehrerbildung. *Journal für LehrerInnenbildung*, 76–84. https://doi.org/10.35468/jlb-02-2020_06

Lim, L., Bannert, M., van der Graaf, J., Molenaar, I., Fan, Y., Kilgour, J., Moore, J., & Gašević, D. (2021). Temporal Assessment of Self-Regulated Learning by Mining Students' Think-Aloud Protocols. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/article/10.3389/fpsyg.2021.749749

Malmberg, J., Järvelä, S., Holappa, J., Haataja, E., Huang, X., & Siipo, A. (2019). Going beyond what is visible: What multichannel data can reveal about interaction in the context of collaborative learning? *Computers in Human Behavior*, *96*, 235–245. https://doi.org/10.1016/j.chb.2018.06.030

Malmberg, J., Järvenoja, H., & Järvelä, S. (2010). Tracing elementary school students' study tactic use in gStudy by examining a strategic and self-regulated learning. *Computers in Human Behavior*, *26*(5), 1034–1042. https://doi.org/10.1016/j.chb.2010.03.004

Matcha, W., Gašević, D., Uzir, N. A., Jovanović, J., & Pardo, A. (2019). Analytics of Learning Strategies: Associations with Academic Performance and Feedback. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 461–470. https://doi.org/10.1145/3303772.3303787

McLaughlin, T., & Yan, Z. (2017). Diverse delivery methods and strong psychological benefits: A review of online formative assessment. *Journal of Computer*

Assisted Learning, *33*(6), 562–574. https://doi.org/10.1111/jcal.12200

Mingorance Estrada, Granda Vera, Rojas Ruiz, & Alemany Arrebola. (2019). Flipped Classroom to Improve University Student Centered Learning and Academic Performance. *Social Sciences*, *8*(11), 315. https://doi.org/10.3390/socsci8110315

Paans, C., Molenaar, I., Segers, E., & Verhoeven, L. (2019). Temporal variation in children's self-regulated hypermedia learning. *Computers in Human Behavior*, *96*, 246–258. https://doi.org/10.1016/j.chb.2018.04.002

Paans, C., Molenaar, I., Segers, E., & Verhoeven, L. (2020). Children's Macro-Level Navigation Patterns in Hypermedia and their Relation with Task Structure and Learning Outcomes. *Frontline Learning Research*, *8*(1), 76–95. https://doi.org/10.14786/flr.v8i1.473

Python (3.8.3) [Software]. (2021). Python Software Foundation. https://python.org

Reimann, P., Markauskaite, L., & Bannert, M. (2014). e-Research and learning theory: What do sequence and process mining methods contribute? *British Journal of Educational Technology*, *45*(3), 528–540. https://doi.org/10.1111/bjet.12146

Schmitt, N. (1996). Uses and abuses of coefficient alpha. *Psychological Assessment*, *8*(4), 350–353. https://doi.org/10.1037/1040-3590.8.4.350

Schneider, R., Sachse, K. A., Schipolowski, S., & Enke, F. (2021). Teaching in Times of COVID-19: The Evaluation of Distance Teaching in Elementary and Secondary Schools in Germany. *Frontiers in Education*, *6*, 702406. https://doi.org/10.3389/feduc.2021.702406

Schoor, C., & Bannert, M. (2012). Exploring regulatory processes during a computer-supported collaborative learning task using process mining. *Computers in Human Behavior*, *28*(4), 1321–1331. https://doi.org/10.1016/j.chb.2012.02.016

Sonnenberg, C., & Bannert, M. (2015). Discovering the Effects of Metacognitive Prompts on the Sequential Structure of SRL-Processes Using Process Mining Techniques. *Journal of Learning Analytics*, *2*(1). https://doi.org/10.18608/jla.2015.21.5

Sonnenberg, C., & Bannert, M. (2016). Evaluating the Impact of Instructional Support Using Data Mining and Process Mining: A Micro-Level Analysis of the Effectiveness of Metacognitive Prompts. *Journal of Educational Data Mining*, *8*(2), 51–83.

Sonnenberg, C., & Bannert, M. (2019). Using Process Mining to examine the sustainability of instructional support: How stable are the effects of metacognitive prompting on self-regulatory behavior? *Computers in Human Behavior*, *96*, 259–272. https://doi.org/10.1016/j.chb.2018.06.003

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Titze, S., Schneeweiss, A., & Lewalter, D. (2021). Die Toolbox Lehrerbildung: Eine Lernplattform für die Professionalisierung von Lehramtsstudierenden im digitalen Zeitalter. *E-Teaching.Org-Artikel*.

Veenman, M. V. J. (2011). Alternative assessment of strategy use with self-report instruments: A discussion. *Metacognition and Learning*, *6*(2), 205–211. https://doi.org/10.1007/s11409-011-9080-x

Winne, P. H. (2013). Learning strategies, study skills, and self-regulated learning in postsecondary education. In *Higher Education: Handbook of Theory and Research* (pp. 377–403). Springer.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In *Handbook of Self-Regulation* (pp. 531–566). Elsevier.