

***Digitalization in Economic and Business Education –
Exploring how technology can support and enhance
educational processes***

Inaugural Dissertation

to Obtain the Academic Degree

of a Doctor in Business Administration

at the University of Mannheim

submitted by

Christian W. Mayer

Master of Science Business Education

Mannheim

Dean

Joachim Lutz

Referent

Professor Dr. Jürgen Seifried, University of Mannheim

Correferent

Professor Dr. Andreas Rausch, University of Mannheim

Day of oral examination

September 25, 2024

“Those who understand, teach.” – Lee Shulman (1986, p. 14)

Abstract

Digitalization is a constant driver of challenges and opportunities in vocational education and training. This thesis addresses key findings from five underlying studies, discussing their limitations, and concludes with practical implications and the relevance of this work. The dissertation deals first with new opportunities and challenges that digitalization brings to economic and business education in vocational education and training, and the implications for teaching, learning, and research.

Five studies are then presented that deal with different research focal points of digitalization in relation to teaching, learning and research. The first study examines the current use of digital tools in commercial vocational schools in times of the corona crisis and concludes that most tools are used to improve rather than change teaching practice and that teachers are likely to be skeptical about whether digital improvements will last after the pandemic. The second study deals with the question of how future teachers can be trained for teaching and learning with ERP systems and emphasizes the need for training already in the university phase of teacher education. The third study reviews eye-tracking research to gain insights into the learning processes in computer-based learning simulations and highlights the potential of eye-tracking as a methodological tool that is currently underrepresented in commercial education research. The fourth study applies the method to analyze visual attention during learning with static and dynamic graphics. It concludes that dynamic graphics support learning better and links the results to visual attention processes. The fifth study evaluates the effectiveness of AI in scoring student responses and shows that AI can support, but not completely replace, human raters.

Limitations of these studies include sample bias, reliance on self-reported data, and the need for more comprehensive skill assessments. Future research should therefore focus on exploring the integration of AI and human judgment as a separate research focus, expanding the use of eye-tracking in complex problem-solving research, and further improving the training of prospective teachers in the use of digital tools. In conclusion, this thesis provides a comprehensive overview of the impact of digitalization on VET and offers suggestions for future research.

Acknowledgements

Diese Arbeit wäre ohne die Unterstützung vieler wunderbarer Menschen nicht möglich gewesen, und ich möchte diese Gelegenheit nutzen, um ihnen zu danken.

Zuallererst möchte ich mich bei meinem Doktorvater, Prof. Dr. Jürgen Seifried, bedanken, dessen Fachwissen und Unterstützung maßgeblich zum Erfolg dieser Arbeit beigetragen haben. Er hat mir viele Freiheiten gelassen und mich mit wertvollen Anregungen unterstützt. Durch meine Arbeit am Lehrstuhl konnte ich an nationalen und internationalen Konferenzen teilnehmen, was mir sowohl beruflich als auch im Austausch mit anderen Wissenschaftler*innen sehr geholfen hat. Für diese Möglichkeiten bin ich sehr dankbar. Ein besonderer Dank geht auch an Prof. Dr. Andreas Rausch, der das Zweitgutachten meiner Dissertation angefertigt hat. Seine Unterstützung während meines gesamten Promotionsstudiums war von großem Wert.

Ein großes Dankeschön gebührt auch dem großartigen Team des Lehrstuhls. Besonders danken möchte ich Rosemarie Pflüger, der guten Seele des Lehrstuhls, die sich einen wohlverdienten, entspannten und erfüllten Ruhestand redlich verdient hat. Außerdem möchte ich mich bei meinen (ehemaligen) Kolleg*innen bedanken, insbesondere bei Alexander Brodsky, Stefanie Birkle, Sophia Gentner, Teresa Giek und Robert Mühldorfer, die mir oft ein Lächeln ins Gesicht zaubern.

Außerdem schätze ich den Austausch mit meinen Kolleginnen und Kollegen aus der Area Wirtschaftspädagogik während der Forschungskolloquien und Mittagspausen sehr. Ich möchte mich auch bei allen studentischen Hilfskräften bedanken, besonders bei meinem VR-Team: Lina Frischmann, Lena Thomas und Pascal Thiry. Ich möchte mich auch bei allen Student*innen aller Fachrichtungen bedanken, die ich unterrichten durfte, den Studienteilnehmern und der Verwaltung, insbesondere der Medientechnik, allen voran Laura Gelb.

Ein großes Danke geht an die "Teach-R"-Community und meinen geschätzten Kollegen Axel Wiepke für seine inspirierende Art. Ein besonderer Dank gilt auch Stephan Leible, dessen Meinung ich sehr schätze und dessen Forschung ich weiterhin gerne verfolgen werde. Ich möchte mich auch bei meinen Forschungspartnern Steffen Brandt, Sabrina Ludwig, Josef Guggemos und Stefanie Findeisen für die hervorragende Zusammenarbeit bedanken.

Ein besonderer Dank richtet sich an Andrea Klein, die mich nicht nur zu einer Promotion ermutigt, sondern mich auch während des gesamten Prozesses mit ihrer herzlichen Art unterstützt hat. Der gegenseitige Austausch gab mir immer den nötigen Schwung. Mein tiefster Dank gilt zudem meiner Familie und meinen Freunden, die mir in stürmischen Zeiten immer eine Zuflucht boten und mir die Abwechslung gaben, die ich brauchte.

Nicht zuletzt geht mein Dank an meine Partnerin Stefanie Spreng. Du hast es nicht nur all die Jahre mit mir ausgehalten, sondern mich auch immer emotional unterstützt und motiviert. Was auch immer für Herausforderungen kommen mögen, ich weiß, dass wir sie gemeinsam meistern werden.

Vielen Dank an euch alle - ohne euch wäre diese Arbeit nicht möglich gewesen!

für Stefanie

Table of Contents

Abstract

Acknowledgements

Table of Contents I

List of Abbreviations..... IV

List of Figures..... VI

List of Tables VIII

List of Appendices IX

1 Introduction..... 1

1.1 The relevance of digitalization in economic and business education..... 1

1.2 Research questions 3

1.3 Structure of this thesis 7

1.4 References 7

2 Conceptual Foundation 12

2.1 Digitization, digitalization and digital transformation in teaching and learning
..... 12

2.2 Teachers' professional knowledge 15

2.3 Analyzing learning processes in technological-rich environments 23

2.4 Summary 26

2.5 References 27

**3 Digitale Unterrichtspraxis an kaufmännischen Schulen in der Corona-Pandemie - Eine
Momentaufnahme..... 35**

3.1 Abstract 35

3.2 Emergency-Remote-Teaching in der Corona-Pandemie..... 36

3.3 Digitale Tools im Unterricht 37

3.4 Methodisches Vorgehen 41

3.5 Empirische Befunde 45

3.6 Diskussion 52

3.7 Literaturverzeichnis..... 54

**4 Looking behind the scenes - Being able to master ERP systems as a goal of vocational
education and training..... 58**

4.1 Relevance of ERP systems for vocational education and training 58

4.2	Theoretical Foundation for Teaching and Learning with ERP systems	60
4.3	Methodology	65
4.4	Findings	66
4.5	Conclusions	68
4.6	References	70
5	Analyzing domain-specific problem-solving processes within authentic computer-based learning and training environments by using eye-tracking: A scoping review	73
5.1	Abstract	73
5.2	Introduction	74
5.3	Theoretical background	76
5.4	Methodology	79
5.5	Results	82
5.6	Discussion	90
5.7	References	94
6	Visual attention while solving economic learning tasks using dynamic, interactive graphs – An eye-tracking study	102
6.1	Abstract	102
6.2	Introduction	102
6.3	Theoretical background	104
6.4	State of research	105
6.5	The present study	107
6.6	Methodology	109
6.7	Results	112
6.8	Limitations	118
6.9	Discussion	118
6.10	References	121
7	Prompt text classifications with transformer models! An exemplary introduction to prompt-based learning with large language models	126
7.1	Abstract	126
7.2	Introduction	127
7.3	Related Work	129
7.4	Methodology	131
7.5	Experiment	136
7.6	Results	138
7.7	Discussion	140

7.8	Conclusion - Reciprocal Perspectives on AI and Human Intelligence in Education.....	142
7.9	References	145
8	Discussion and Research Outlook	152
8.1	Summary	152
8.2	Limitations and Future Research.....	156
8.3	Practical Implications and Conclusion	158
8.4	References	160
	Complete List of References.....	162
	Appendix	193
	Declaration in Lieu of Oath.....	203
	Doctoral Study Program.....	205
	Curriculum Vitae	206

List of Abbreviations

AES	Automated Essay Scoring
AI	Artificial Intelligence
AIED	Artificial Intelligence in Education
AOI	Area of Interest
ASCOT	Technology-based Assessment of Skills and Competences in VET
BERT	Bidirectional Encoder Representations from Transformers
BIBB	The Federal Institute for Vocational Education and Training (Bundesinstitut für Berufliche Bildung)
CEDEFOP	European Center for the Development of Vocational Training
CFT	Competence Framework for Teachers
CK	Content Knowledge
DESI	Digital Economy and Society Index
DigCompEdu	European Framework for the Digital Competence of Educators
DK	Digital Knowledge
DPACK	Digitality Related Pedagogical Content Knowledge
EDA	Electrodermal Activity
ERP	Enterprise Resource Planning
FACS	Facial Action Coding System
GPT	Generative Pre-trained Transformer
ICILS	International Computer and Information Literacy Study
ICT	Information and Communications Technology
IT	Information Technology
K–12	From Kindergarten to 12th Grade
KMK	Standing Conference of the Ministers of Education and Cultural Affairs of the Länder in the Federal Republic of Germany (Kultusministerkonferenz)

NLP	Natural Language Processing
OECD	Organization of Economic Co-operation and Development
OER	Open Educational Resource
PCK	Pedagogical Content Knowledge
PK	Pedagogical Knowledge
PLM	Pre-Trained Language Models
PNr.	Problem Number
SDG	Sustainable Development Goals
SME	Small and Medium Enterprise
STEM	Science, Technology, Engineering and Mathematics
TCK	Technological Content Knowledge
TK	Technological Knowledge
TPK	Technological Pedagogical Knowledge
TPACK	Technological Pedagogical Content Knowledge
UNESCO	United Nations Educational Scientific and Cultural Organization
VET	Vocational Education and Training
VR	Virtual Reality

List of Figures

Figure 1-1: Key research foci and related studies	4
Figure 2-1: Integrative PCK model, adapted from Shulman (1986)	16
Figure 2-2: Transformative Model of teacher knowledge, adapted from Fernandez (2014; Grossmann 1990)	17
Figure 2-3: TPACK model, adapted from Mishra (2019), Mishra & Koehler (2006). Original Source: tpack.org	18
Figure 2-4: The DPACK model, adapted from Thyssen (2023) including a sociocultural knowledge sphere	19
Figure 2-5: The UNESCO ICT CFT in its third version (UNESCO, 2018, p. 10).....	22
Figure 2-6: The DigCompEdu Framework Areas and Scope for the Digital Competence of Educators (Redecker, 2017, p. 15)	23
Figure 3-1: Stufen des SAMR-Modells (adaptiert von Puentedura, 2006)	40
Figure 3-2: Nutzung von digitalen Tools (eigene Darstellung).....	47
Figure 3-3: Zustimmung bzw. Ablehnung zu digitalen Tools der Enhancement Ebene (eigene Darstellung).....	49
Figure 3-4: Zustimmung bzw. Ablehnung zu digitalen Tools der Transformation Ebene (eigene Darstellung).....	49
Figure 4-1: Curriculum-Instruction-Assessment (CIA) triad in the style of Pellegrino (Achtenhagen, 2012; 2002, 2006).....	60
Figure 4-2: Sales process (simplified) in the style of SAP4school material (<i>SAP4school IUS</i> , 2023).....	62
Figure 4-3: Technological Pedagogical Content Knowledge (Mishra & Koehler, 2006). Source: available at http://tpack.org (TPACK.org, 2023)	64
Figure 5-1: The scope of this review (in bold) is illustrated within a continuum between cross-sectoral and domain-specific problem-solving simulations	75
Figure 5-2: PRISMA flow chart	80

Figure 5-3: Included studies per 5-years interval	83
Figure 6-1: Learning task with a dynamic, interactive graph that allows manipulation to answer several single-choice tasks (not in the picture).....	108
Figure 6-2: The experimental procedure	111
Figure 6-3: The predicted mean score for dynamic, interactive graphics exercises by average fixation duration (both standardized)	115

List of Tables

Table 1-1: Overview of the studies in this dissertation	6
Table 3-1: Ausschnitt aus dem Kodierleitfaden zur Kategorie Unterrichtsgestaltung.....	45
Table 3-2: Erwartete langfristige Auswirkungen der pandemiebedingten Digitalisierung des Unterrichts auf die postpandemische Unterrichtspraxis (Anzahl der jeweiligen (Sub-)Codes)	50
Table 5-1: Key features, inclusion criteria, and exclusion examples	81
Table 5-2: Overview of domains	83
Table 5-3: Overview of eye-tracking setup, calculated measures, analysis techniques, and other collected measures	84
Table 6-1: Performance Scores of Participants	113
Table 6-2: Repeated measurement ANOVA of eye-movement measures.....	114
Table 6-3: Post hoc analysis for Fixation Duration (ms)	114
Table 6-4: Regression results using mean performance as the dependent variable	116
Table 6-5: Regression results using mean performance as the dependent variable	117
Table 7-1: Overview of the models for the experiment	137
Table 7-2: Performance per model.....	138
Table 7-3: Confusion matrix per model	139
Table 7-4: Performance metrics for fine-tuning.....	139
Table 7-5: Confusion matrix for fine-tuning.....	140

List of Appendices

Appendix A: Study 3 (Chapter 5). Analysing domain-specific problem-solving processes within authentic computer-based learning and training environments by using eye-tracking: A scoping review	193
Appendix B: Study 4 (Chapter 6). Visual attention while solving economic learning tasks using dynamic, interactive graphs – An eye-tracking study	200
Appendix C: Study 5 (Chapter 7). Prompt text classifications with transformer models! An exemplary introduction to prompt-based learning with large language models	202

1 Introduction

This thesis commences with an overview on the relevance of digitalization in economic and business education (section 1.1). This is followed by a brief description of the research questions on which this thesis is based (section 1.2) and an overview of the structure of this thesis (section 1.3).

1.1 The relevance of digitalization in economic and business education

Digitalization is one of several drivers of change in today's world. The digital transformation is having a profound impact on the economy and society and is changing the way we work, live, and learn. The European Commission has declared a Digital Decade and is constantly monitoring the progress of its Member States (European Commission, 2023) in terms of digital skills, digital infrastructure, digitization of businesses, and digitization of public services. According to the Digital Economy and Society Index (DESI), which has become part of the Digital Decade reporting, an indicator for the digital transformation of companies shows the underlying level of digital intensity in small and medium-sized enterprises (SMEs), which is constantly increasing. Thus, the requirements for the training and further education of employees' digital skills are increasingly changing. Additionally, it has long been known that computers are increasingly able to perform tasks that were previously carried out exclusively by humans (Brynjolfsson & McAfee, 2016), and the latest advances in artificial intelligence (AI) will further influence the labor market (Frank et al., 2019). Office workplaces are increasingly using software for organizational processes and decision-making (Billett, 2021), and workplace (digital) skills are linked to tasks and roles that are facilitated and automated by electronic means, while AI is expected to become the next co-worker or co-teacher in an era of co-intelligence (Mollick, 2024).

This megatrend also applies to the education sector. Germany in particular is considered a laggard in digitalization (in the education sector) by international comparison and has recently responded to the lack of digitalization in the education sector with national policy plans (Zancajo et al., 2022). In general, as the current national education report shows, digitalization can be seen as a consistent and cross-sectoral driver of change in almost all areas of the education system, from early childhood through K–12 to adult education and lifelong learning (Autor:innengruppe Bildungsberichterstattung, 2024). In particular, digitalization directly affects vocational education and training (VET) domains with its close dependencies to changes in the labor market and professional requirements. VET must take account of the changing requirements of the labor market. In Germany, this applies in particular to the dual VET system with two learning venues that are embedded in the overall German VET system (Deissinger, 2015).

Changes to the occupational profile at the company learning location must be taken into account not only at the company level of dual training, but also at the vocational school learning site, e.g. in order to train digital skills that are not yet standard in every company but are relevant for the fulfillment of occupational characteristics and training requirements, as well as to promote the further placement ability (i.e., employability) of trainees (Seifried et al., 2019; Zutavern & Seifried, 2022). In the field of economic and business education, for example, the use of enterprise resource planning (ERP) systems (Knigge et al., 2017; Mayer, 2022) can be seen as such an adaptation to the current changes in professional requirements and can prove beneficial in promoting business process knowledge among trainees in addition to motivation and interest (Spener et al., 2019; Spener & Schumann, 2021). One such example for industrial-technical vocational training could be the implementation of learning factories to promote subject-specific skills in authentic industrial work processes (Roll & Ifenthaler, 2020, 2021). Similarly, a property management system adopted from the hotel industry is being introduced into teacher training for nutrition and home economics (Klatt et al., 2023). However, it seems clear that (prospective) teachers need to be trained to teach with such complex digital teaching and learning arrangements.

The trend toward increased digitization in education, accelerated by the Coronavirus Pandemic, will continue (OECD & The World Bank, 2022). However, meaningful learning with digital tools depends on school leadership, technological infrastructure, and teachers' willingness to use these tools (Delcker & Ifenthaler, 2021). Introducing digital tools in higher education for preservice teachers could increase their readiness if they believe it will improve learning (Petko, 2012). Prospective teachers are expected to teach with and about ICT (Frailon et al., 2019), and they need to be prepared for digitization (Dillenbourg, 2013; Seufert et al., 2021). This includes the integration of modern technologies in VET, not only for teaching and learning, but also for research. VET initiatives like 'Technology-based Assessment of Skills and Competences in VET' (ASCOT) provides researchers and practitioners with authentic computer-based learning environments and simulations (Breckwoldt et al., 2014) to develop domain-specific problem-solving skills (Beck et al., 2016; Rausch & Wuttke, 2016). Much of the past research in the field of vocational education and training has focused on outcomes, such as competency assessment and learning performance, rather than on the processes that lead to these outcomes (Abele, 2018). It is therefore important to emphasize the methodological possibilities of technological advances in the collection and analysis of learning process data, such as eye-tracking or logfile data, in order to gain deeper insights. Computer-based simulations enable individualized learning through the analysis of logfiles (Ludwig et al., 2024) and insights into learners processes by analyzing eye-tracking data (Lee et al., 2019). Additionally, with the advent of

AI, future teachers need to engage with AI tools such as automated essay scoring systems (Celik et al., 2022; Ludwig et al., 2021), and develop AI readiness (Wang et al., 2023) while critically evaluating AI integration capabilities (Lameras & Arnab, 2021).

In summary, this thesis explores digitization in business education by examining how technology can improve educational processes. In view of the developments outlined so far, the relevance of digitalization for VET (Wuttke et al., 2020) and economic and business education is considerable. First, the changing occupational characteristics and demands of the labor market are likely to have an impact on the business education system, and therefore teachers need to be prepared to teach with digital tools to foster digital skills among learners. In this work, the current use of digital tools in commercial schools is first analyzed (Mayer, Gentner, et al., 2023) and further exemplary shown and discussed by integrating a seminar on teaching and learning with ERP systems in teacher training curriculum (Mayer & Seifried, 2024). Second, individualized learning can be promoted through digital means using complex teaching and learning arrangements such as computer-based simulations. In order to gain a deeper understanding of individual learning behavior, the use of methodological approaches such as log file analysis and eye-tracking is crucial to understand individual learning processes. Therefore, this work provides a systematic review of research to illustrate the methodological analysis of individual learning, focusing on eye-tracking studies in vocational training simulations to understand these learning processes (Mayer, Rausch, et al., 2023), and consequently conducts a first eye-tracking study in economic education (Mayer et al., under review). Third, despite the current challenges of digitalization, new technologies such as AI need to be critically evaluated to prepare (prospective) teachers for future opportunities. Finally, in line with the advances of AI as a possible future co-teacher, an automated text scoring procedure is critically evaluated using a domain-specific scoring task from the field of economics and business education (Mayer et al., 2022) to determine the future potential of AI in Education (AIED).

1.2 Research questions

As outlined so far, digitalization is a constant driver and poses new challenges for economic and business education. To meet these challenges, this dissertation focuses on the professional development of (prospective) teachers and analyses *how technology can support and improve educational processes*. To answer this question, three main areas of research can be distinguished. Figure 1-1 illustrates the research foci and studies in this thesis. Table 1-1 provides an overview over these studies.

The first research focus refers to the current use of digital tools by teachers in commercial vocational schools and raises the question of *what the current state of the art is for digital tools in commercial vocational schools* (study 1).

The COVID-19 pandemic led to school closures and a rapid switch to distance learning, so-called emergency remote learning (Dilger, 2021). Study 1 focuses on the current use of digital tools by teachers at commercial vocational schools during the pandemic, based on the responses of 305 teachers. The study raises the question of how digital tools are being used in vocational schools. And what expectations teachers have that digital tools will continue to supplement face-to-face teaching after the pandemic.

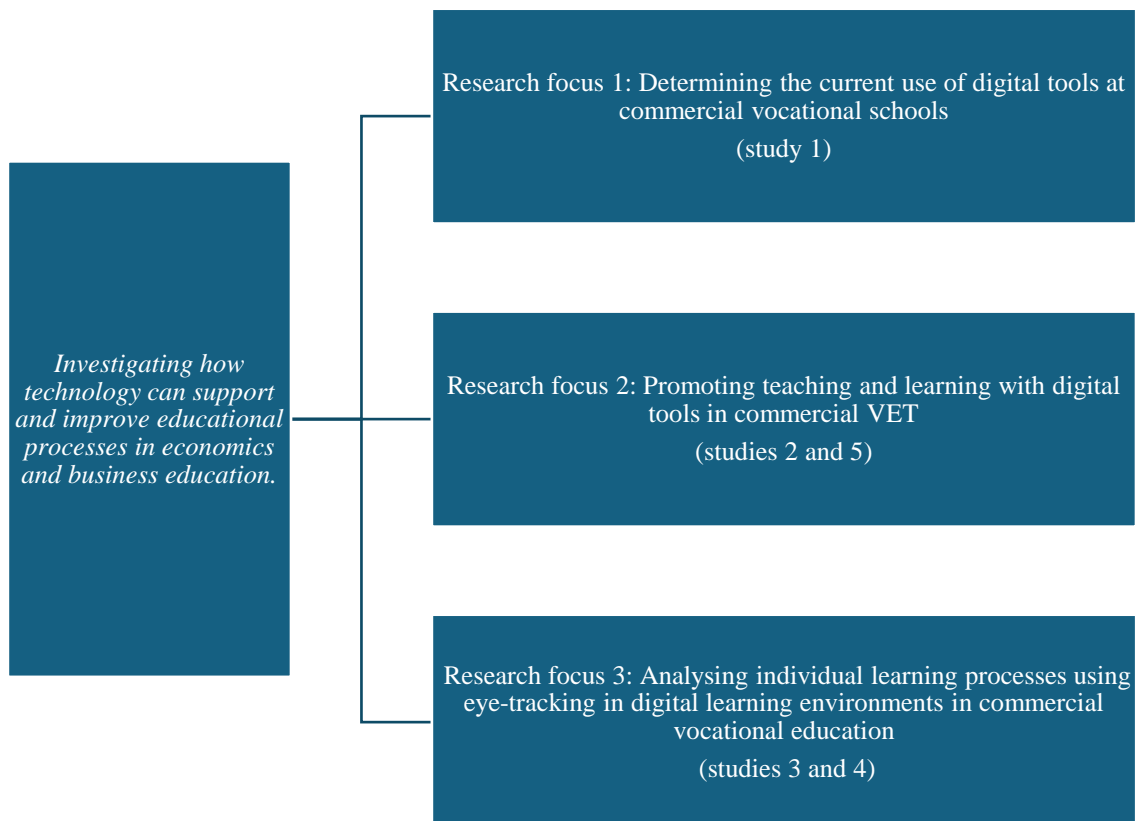


Figure 1-1: Key research foci and related studies.

The second research focus relates to teaching and learning with digital tools VET and explores the question of *how (prospective) teachers can be explicitly trained and supported through the use of sophisticated digital tools* by implementing a course in prospective teacher education for teaching and learning with ERP systems (study 2), as well as the application of AI prompting to assess whether a large language model can automatically and reliably assess students' responses to support researchers and teachers (study 5).

Study 2 discusses the benefits of integrating ERP systems into the training of apprentices, both in vocational schools and in the workplace. To ensure high quality ERP training, prospective teachers and in-company trainers must understand the role of ERP systems in business processes. In order to promote teaching and learning with ERP systems, a seminar was designed for future in-company trainers and vocational teachers. The study evaluates a teaching-learning concept for ERP systems among 26 master students in economic and business education using a pre-post questionnaire on technological pedagogical content knowledge (TPACK; Mishra & Koehler, 2006). Study 5 investigates the potential of automated text classification using prompt-based learning approaches with transformer models for a domain-specific task. Prompt-based learning, which requires little or no training examples, allows the use of artificial intelligence in education (Popenici & Kerr, 2017) without sophisticated programming skills or extensive labeled data. The study compares zero-shot and few-shot classification approaches with a fine-tuned language model and human ratings. The dataset consists of 2,088 email responses labeled for professional communication style.

The third research focus relates to individual learning processes and *how the use of unobtrusive measuring instruments, such as eye-tracking, can provide insights into learning processes in digital learning environments*. To this end, the current state of eye-tracking research in computer-based simulations in VET was analyzed (study 3) and consequently, a first eye-tracking study was carried out in a computer-based learning environment for economic education (study 4).

Study 3 examines the use of eye-tracking to analyze complex problem-solving processes in authentic computer-based learning and training environments, particularly in VET. The scoping review according to the PRISMA scheme (Liberati et al., 2009) identifies 12 relevant studies and reports on the current state of research in this area. Study 4 then uses eye-tracking to investigate visual attention and how visualizations, particularly price-quantity diagrams, can enhance learning in economics in a computer-based learning environment. While the contributions of eye-tracking research to multimedia learning is extensive (Alemdag & Cagiltay, 2018), there are few studies that deal with static and dynamic, interactive graphs in economic education. In this study, the eye movements of 31 economics students were analyzed with regard to their gaze behavior when solving economic single-choice learning tasks with the help of static and dynamic graphs.

Table 1-1: Overview of the studies in this dissertation.

Study 1	
Reference	Mayer, C. W., Gentner, S., & Seifried, J. (2023). Digitale Unterrichtspraxis an kaufmännischen Schulen in der Corona-Pandemie - Eine Momentaufnahme [Digital teaching practice at commercial schools during the corona pandemic - a snapshot]. <i>Zeitschrift für Berufs- und Wirtschaftspädagogik</i> , 119(3), 531–552. https://doi.org/10.25162/zbw-2023-0021
Research foci	1. Determining the current use of digital tools at commercial vocational schools
Research approach	Questionnaire
Analysis	Quantitative descriptive and qualitative content analysis (mixed methods)
Sample	305 teachers at commercial vocational schools
Study 2	
Reference	Mayer, C. & Seifried, J. (2024). Looking behind the scenes: being able to master ERP systems as a goal of vocational education and training. In Cedefop, OECD (Eds.), <i>Apprenticeships and the digital transition: modernising apprenticeships to meet digital skill needs</i> (pp. 118–127). Publications Office of the European Union. https://data.europa.eu/doi/10.2801/074640
Research foci	2. Promoting teaching and learning with digital tools in commercial VET
Research approach	Learning diaries and questionnaires
Analysis	Quantitative analysis and qualitative content analysis (mixed methods)
Sample	26 students in economic and business education
Study 3	
Reference	Mayer, C. W., Rausch, A. & Seifried, J. (2023). Analysing domain-specific problem-solving processes within authentic computer-based learning and training environments by using eye-tracking: a scoping review. <i>Empirical Research in Vocational Education and Training</i> , 15:2, 1–27. https://doi.org/10.1186/s40461-023-00140-2
Research foci	3. Analyzing individual learning processes in digital learning environments
Research approach	Literature review
Analysis	Scoping review
Sample	12 studies
Study 4	
Reference	Mayer, C. W., Findeisen, S., Guggemos, J. & Seifried, J. (under review). Visual attention while solving economic learning tasks using dynamic graphs – An eye-tracking study. <i>Journal of Economic Education</i> .
Research foci	3. Analyzing individual learning processes in digital learning environments
Research approach	Eye-Tracking and questionnaires
Analysis	Quantitative analysis
Sample	31 students in economic and business education
Study 5	
Reference	Mayer, C. W. F., Ludwig, S. & Brandt, S. (2023). Prompt text classifications with transformer models! An exemplary introduction to prompt-based learning with large language models. <i>Journal of Research on Technology in Education: JRTE</i> , 55(1), 125–141. https://doi.org/10.1080/15391523.2022.2142872
Research foci	2. Promoting teaching and learning with digital tools in commercial VET
Research approach	Machine learning, few-shot learning
Analysis	Quantitative analysis
Sample	2,088 email responses to a domain-specific problem-solving task

In summary, the results of the five studies may serve as a guide for answering the question of the impact of digitalization in economic and business education. First, the thesis identifies the

current use of digital tools in vocational schools. Second, the thesis proposes a training course for master's students on learning and teaching with an advanced learning environment (ERP system) and further evaluates technological innovations, such as large language models, for a simple domain-specific assessment task. Third, the studies analyze individual learning processes in technology-rich environments using eye-tracking as a methodological approach to examine problem-solving processes in authentic learning environments. After conducting a systematic review, it applies the methodological approach to understand learning processes and improve instructional design in computer-based learning environments.

1.3 Structure of this thesis

This inaugural dissertation comprises five studies and is structured into eight chapters. The thesis commences with an introduction to the relevance of digitalization in economics and business education (section 1.1) and the research questions that are examined in the five studies that comprise this thesis (section 1.2). Subsequently, an overview of the structure of the thesis is provided (section 1.3). The conceptual foundations underlying the five studies are discussed in Chapter 2. The chapter commences with a concise definition of the term "digitalization" (section 2.1), and distinctions between digital transformation, digitalization, and digitization, and a brief look at past and current hypes. Subsequently, the chapter examines the professional knowledge of teachers in terms of technological, pedagogical, and content knowledge (section 2.2). This chapter then introduces eye-tracking as a methodological approach to understanding learning processes in technology-rich environments and demonstrates the connection to TPACK as it relates to teachers' professional knowledge of learning and instructional design (section 2.3). The following five chapters comprise the five studies that make up this thesis: Chapters 3, 4, 5, 6, and 7. The following chapter 8 summarizes the main findings of the five studies, accompanied by an analysis of their limitations and a discussion of the implications for future research.

1.4 References

- Abele, S. (2018). Diagnostic Problem-Solving Process in Professional Contexts: Theory and Empirical Investigation in the Context of Car Mechatronics Using Computer-Generated Log-Files. *Vocations and Learning*, *11*(1), 133–159. <https://doi.org/10.1007/s12186-017-9183-x>
- Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, *125*, 413–428. <https://doi.org/10.1016/j.compedu.2018.06.023>

- Autor:innengruppe Bildungsberichterstattung. (2024). *Bildung in Deutschland 2024. Ein indikator-gestützter Bericht mit einer Analyse zu beruflicher Bildung [Education in Germany 2024. An indicator-based report with an analysis of vocational education and training]*. wbv.
- Beck, K., Landenberger, M., & Oser, F. (Eds.). (2016). *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der BMBF-Förderinitiative ASCOT [Technology-based skills assessment in vocational education and training: results from the BMBF funding initiative ASCOT]*. W. Bertelsmann.
- Billett, S. (2021). Mediating worklife learning and the digitalisation of work. *British Journal of Educational Technology*, 52(4), 1580–1593. <https://doi.org/10.1111/bjet.13115>
- Breckwoldt, J., Gruber, H., & Wittmann, A. (2014). Simulation Learning. In S. Billett, C. Harteis, & H. Gruber (Eds.), *International Handbook of Research in Professional and Practice-based Learning* (pp. 673–698). Springer. https://doi.org/10.1007/978-94-017-8902-8_25
- Brynjolfsson, E., & McAfee, A. (2016). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton.
- Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The Promises and Challenges of Artificial Intelligence for Teachers: A Systematic Review of Research. *TechTrends*, 66(4), 616–630. <https://doi.org/10.1007/s11528-022-00715-y>
- Deissinger, T. (2015). The German dual vocational education and training system as ‘good practice’? *Local Economy: The Journal of the Local Economy Policy Unit*, 30(5), 557–567. <https://doi.org/10.1177/0269094215589311>
- Delcker, J., & Ifenthaler, D. (2021). Teachers’ perspective on school development at German vocational schools during the Covid-19 pandemic. *Technology, Pedagogy and Education*, 30(1), 125–139. <https://doi.org/10.1080/1475939X.2020.1857826>
- Dilger, B. (2021). Lernen aus den Erfahrungen mit Emergency-Remote-Lehre aus Studierenden- und Dozierenden-Perspektive [Learning from experiences with emergency remote teaching from the perspective of students and lecturers]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117(1), 3–13. <https://doi.org/10.25162/zbw-2021-0001>
- Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, 69, 485–492. <https://doi.org/10.1016/j.compedu.2013.04.013>
- European Commission (Ed.). (2023). *Digital Economy and Society Index (DESI) 2023: Integration of digital technology*. <https://digital-decade-desi.digital-strategy.ec.europa.eu/datasets/desi/charts>
- Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2019). *Preparing for life in a digital world: IEA international computer and information literacy study 2018 international report*. International Association for the Evaluation of Educational Achievement (IEA). <https://www.iea.nl/studies/iea/icils/2018>
- Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences*, 116(14), 6531–6539. <https://doi.org/10.1073/pnas.1900949116>
- Klatt, C., Härtel, S., & Markert, J. (2023). Digitalisierung im Berufsfeld Ernährung und Hauswirtschaft – Einsatz eines Property-Management-Systems aus der Hotellerie in der Lehramtsbildung für berufsbildende Schulen [Digitization in the Vocational Field of Nutrition and Home Economics—Implementation of a property management system

- adapted from the hotel industry in teacher education for vocational schools]. *bwp@ Berufs- und Wirtschaftspädagogik – online*, 45, 1–20.
- Knigge, M., Prifti, L., Kienegger, H., & Krcmar, H. (2017). Teaching enterprise organization and enterprise resource planning systems in schools: Playing a serious game with pupils. *2017 IEEE Global Engineering Education Conference (EDUCON)*, 486–495. <https://doi.org/10.1109/EDUCON.2017.7942891>
- Lameras, P., & Arnab, S. (2021). Power to the Teachers: An Exploratory Review on Artificial Intelligence in Education. *Information*, 13(1), 14. <https://doi.org/10.3390/info13010014>
- Lee, J. Y., Donkers, J., Jarodzka, H., & Merrienboer, J. van. (2019). How prior knowledge affects problem-solving performance in a medical simulation game: Using game-logs and eye-tracking. *Computers in Human Behavior*, 99, 268–277. <https://doi.org/10.1016/j.chb.2019.05.035>
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ (Clinical Research Ed.)*, 339, b2700. <https://doi.org/10.1136/bmj.b2700>
- Ludwig, S., Mayer, C., Hansen, C., Eilers, K., & Brandt, S. (2021). Automated essay scoring using transformer models. *Psych*, 3(4), 897–915.
- Ludwig, S., Rausch, A., Deutscher, V., & Seifried, J. (2024). Predicting problem-solving success in an office simulation applying N-grams and a random forest to behavioral process data. *Computers & Education*, 218, 105093. <https://doi.org/10.1016/j.compedu.2024.105093>
- Mayer, C. W. (2022). Lehren und Lernen mit Enterprise Resource Planning (ERP) Systemen—Typische Bearbeitungsprobleme als Grundlage der Lernprozessgestaltung [Teaching and learning with Enterprise Resource Planning (ERP) systems Typical processing problems as a basis for designing the learning process]. *Bwp@ Berufs- Und Wirtschaftspädagogik – Online*, 43, 1–17.
- Mayer, C. W., Findeisen, S., Guggemos, J., & Seifried, J. (under review). Visual attention while solving economic learning tasks using dynamic graphs – An eye-tracking study. *Journal of Economic Education*.
- Mayer, C. W., Gentner, S., & Seifried, J. (2023). Digitale Unterrichtspraxis an kaufmännischen Schulen in der Corona-Pandemie—Eine Momentaufnahme [Digital teaching practice at commercial schools during the corona pandemic - a snapshot]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 119(3), 531–552. <https://doi.org/10.25162/zbw-2023-0021>
- Mayer, C. W., Ludwig, S., & Brandt, S. (2022). Prompt text classifications with transformer models! An exemplary introduction to prompt-based learning with large language models. *Journal of Research on Technology in Education*, 55(1), 1–17. <https://doi.org/10.1080/15391523.2022.2142872>
- Mayer, C. W., Rausch, A., & Seifried, J. (2023). Analysing domain-specific problem-solving processes within authentic computer-based learning and training environments by using eye-tracking: A scoping review. *Empirical Research in Vocational Education and Training*, 15(1), 2. <https://doi.org/10.1186/s40461-023-00140-2>
- Mayer, C. W., & Seifried, J. (2024). Looking behind the scenes—Being able to master ERP systems as a goal of vocational education and training. In Cedefop & OECD (Eds.), *Apprenticeships and the digital transition* (pp. 118–127). Publications Office.

- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record: The Voice of Scholarship in Education*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Mollick, E. (2024). *Co-intelligence: Living and working with AI*. W H Allen.
- OECD & The World Bank. (2022). *How Learning Continued during the COVID-19 Pandemic: Global Lessons from Initiatives to Support Learners and Teachers* (S. Vincent-Lancrin, C. Cobo Romani, & F. Reimers, Eds.). OECD Publishing. <https://doi.org/10.1787/bbeca162-en>
- Petko, D. (2012). Teachers' pedagogical beliefs and their use of digital media in classrooms: Sharpening the focus of the 'will, skill, tool' model and integrating teachers' constructivist orientations. *Computers & Education*, 58(4), 1351–1359. <https://doi.org/10.1016/j.compedu.2011.12.013>
- Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12, 22. <https://doi.org/10.1186/s41039-017-0062-8>
- Rausch, A., & Wuttke, E. (2016). Development of a Multi-Faceted Model of Domain-Specific Problem-Solving Competence and its Acceptance by Different Stakeholders in the Business Domain. *Unterrichtswissenschaft*, 44(2), 169–184.
- Roll, M., & Ifenthaler, D. (2020). The Impact of Learning Factories on Multidisciplinary Digital Competencies. In E. Wuttke, J. Seifried, & H. Niegemann (Eds.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (pp. 23–38). Barbara Budrich. <https://doi.org/10.2307/j.ctv18dvv1c>
- Roll, M., & Ifenthaler, D. (2021). Learning Factories 4.0 in technical vocational schools: Can they foster competence development? *Empirical Research in Vocational Education and Training*, 13(1), 20. <https://doi.org/10.1186/s40461-021-00124-0>
- Seifried, J., Beck, K., Ertelt, B.-J., & Frey, A. (Eds.). (2019). *Beruf, Beruflichkeit, Employability [occupation, professionalism, employability]*. wbv.
- Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior*, 115, 106552. <https://doi.org/10.1016/j.chb.2020.106552>
- Spener, C., Häuber, G., Horlacher, T., & Schumann, S. (2019). Der berufsschulische Einsatz von ERP-Systemen in der Ausbildung von Industriekaufleuten [The vocational use of ERP systems in the training of industrial clerks]. In K. Wilbers (Ed.), *Digitale Transformation kaufmännischer Bildung: Ausbildung in Industrie und Handel hinterfragt* (pp. 195–216). epubli. https://www.pedocs.de/frontdoor.php?source_opus=17968
- Spener, C., & Schumann, S. (2021). Wissensseffekte des ERP-Einsatzes in der kaufmännischen Berufsschule [Knowledge effects of ERP use in commercial vocational schools]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117(3), 395–430. <https://doi.org/10.25162/zbw-2021-0018>
- Wang, X., Li, L., Tan, S. C., Yang, L., & Lei, J. (2023). Preparing for AI-enhanced education: Conceptualizing and empirically examining teachers' AI readiness. *Computers in Human Behavior*, 146, 107798. <https://doi.org/10.1016/j.chb.2023.107798>
- Wuttke, E., Seifried, J., & Niegemann, H. M. (Eds.). (2020). *Vocational education and training in the age of digitization: Challenges and opportunities*. Barbara Budrich.

- Zancajo, A., Verger, A., & Bolea, P. (2022). Digitalization and beyond: The effects of Covid-19 on post-pandemic educational policy and delivery in Europe. *Policy and Society*, *41*(1), 111–128. <https://doi.org/10.1093/polsoc/puab016>
- Zutavern, S., & Seifried, J. (2022). Vocational Education and Training in Germany: Benefits and Drawbacks of the Dual Approach as Preparation for Professional Employment. In C. Harteis, D. Gijbels, & E. Kyndt (Eds.), *Research Approaches on Workplace Learning* (pp. 347–365). Springer. https://doi.org/10.1007/978-3-030-89582-2_16

2 Conceptual Foundation

The conceptual basis for the five studies in this thesis is explained below. First, this chapter begins with a brief definition of digitalization and the distinction between digitization, digitalization and digital transformation and provides examples of digitalization for VET. A brief look at past and current hypes and recent developments of digitalization in education is also provided (section 2.1). Second, the professional knowledge of teachers in dealing with digitalization is discussed in section 2.2, by describing the technological pedagogical content knowledge (TPACK) model (Mishra & Koehler, 2006), starting with its predecessor (Shulman, 1986) as well as concluding its potential successor, the DPACK model (Huwer et al., 2019a). In addition, the latest digital competence frameworks for educators will be presented as current policy developments to promote teachers' professional knowledge with regard to digitalization in education. Third, methodological advances in VET research through digitization are discussed using the example of the use of eye-tracking analysis in technology-enhanced environments as a method for measuring learning processes in VET, thus concluding the theoretical framework on which this thesis is based (section 2.3).

2.1 Digitization, digitalization and digital transformation in teaching and learning

The terms *digitization*, *digitalization* and *digital transformation* are often used interchangeably (Frenzel-Piasentin et al., 2021; Gradillas & Thomas, 2023; Parviainen et al., 2017), although these terms have different meanings. Digitization refers to the conversion of previously analogue data into a digital form, while digitalization refers to an entire process that changes from analogue to digital forms, including the use of digital technologies to change a business model. According to Legner et al., the term digitalization refers to “the manifold sociotechnical phenomena and processes of adopting and using these technologies in broader individual, organizational, and societal contexts” (2017, p. 301). The digital transformation expands this perspective further by affecting the entire economy and society and encompassing completely new business models (e.g., Mugge et al., 2020; Vrana & Singh, 2021). This thesis follows the definition of digitalization as a sociotechnical phenomenon as it is described by Legner et al. (2017), and discusses the digitalization of learning in VET, including the individual learning context, the organizational context of the two learning venues in the German vocational system, as well as the societal context of digitalization in vocational education.

VET has been profoundly impacted by the changing digital landscape (Wuttke et al., 2020). The VET sector seems to benefit more, probably because educational technology is primarily

aimed at improving the efficiency of learning processes and promoting flexibility, while general education schools pursue broader educational goals (Niegemann, 2020). According to Euler and Wilbers (2020), VET is more challenged than other areas of education because it faces the challenge of preparing trainees and vocational students for a future that is only vaguely foreseeable. Digital technologies influence VET in terms of (1) new learning tools, (2) new work tools and (3) as universal tools that influence everyday life and learning prerequisites. Furthermore, it is clear that the process of digitalization of learning also has an impact on VET research (see section 2.3; Euler & Wilbers, 2020).

Firstly, digital learning tools change the way we learn. They enable the creation of new organizational learning arrangements, such as the integration of technology-supported learning phases into face-to-face teaching (Euler & Wilbers, 2020). In addition, digital tools help to analyze individual learning processes and adapt learning content to the needs of learners (e.g., Rausch et al., 2021). Teachers also benefit from digital tools, for example by gaining an overview of individual learning progress by visualizing learners' behavior in dashboards (e.g., Gorshid et al., 2022).

Secondly, new work equipment is constantly being introduced into the workplace. There is solid empirical evidence of the link between the digitalization of work with such tools and an increase in productivity in EU countries (Gal et al., 2019). Changing work and business processes become part of the learning content in VET. More precisely, the learning content includes the digital technologies themselves or the technology-supported work and business processes that are relevant to a vocational field of application (Euler & Wilbers, 2020). In the commercial area, ERP systems change the way we work and support employees with adaptive support systems, so-called Electronic Performance Support Systems, when help is needed (e.g., Leiß et al., 2022). Job profiles and qualification requirements are changing, leading to the introduction of digital tools such as ERP systems in commercial vocational schools and training (e.g., Mayer & Seifried, 2022; Spener & Schumann, 2021).

Thirdly, digital tools also function as universal tools. Nowadays, information about almost everything and communication with peers and colleagues are generally accessible anytime and anywhere. The consumption of digital products is becoming a fundamental need of the younger generation (Euler & Wilbers, 2020), which is often referred to as 'digital natives'. The challenges of teaching digital natives is a continuing academic debate (Evans & Robertson, 2020). According to Prensky (2001a, 2001b), digitally literate learners have grown up surrounded by technology and therefore need a fundamental change in the way they are taught. However, the

digital native stereotype does not accurately represent how young people perceive and use technology. The digital native debate has been perceived as an ‘academic panic’ rather than an empirical observation, where generation has been shown to be one of many significant variables (e.g., usage, experience, gender, education) that predict advanced Internet use, suggesting that any generation of learners can become digital natives (Bennett et al., 2008; Helsper & Eynon, 2010). Instead, educators' decisions should be guided by the affordances of technology and the digital literacies needed to use it effectively (E. E. Smith et al., 2020). For example, today's advances in translation tools can foster language learning in everyday life and at school (e.g., Urlaub & Dessen, 2022). Chatbots can support learning processes and improve learning outcomes by acting as individual tutors anytime, anywhere (Hwang & Chang, 2021). Despite the technological advancements that have occurred in recent years, the pedagogically meaningful integration of educational technology remains a persistent challenge, as evidenced by historical precedent.

Historically, the trend towards digitalization in education is nothing new. Educational technology has come a long way in the last century (for a brief overview, see Niegemann, 2020). Niegemann's summary begins in 1913 with a quote by Thomas Edison, who believed that the invention of the motion picture would completely change the school system, resulting in the demise of books (Niegemann, 2020; F. J. Smith, 1913). Motion pictures and audio records were soon replaced by the radio in the 1930s and television broadcast in the 1950s (Molenda, 2022). From programmed or assisted instruction in the 1960s, to teletext technology in the 1980s, to the increased use of computers and the invention of the World Wide Web in 1990, technological innovations were always expected to transform education (Reich, 2020). In recent decades, there has been hype and enthusiasm for various technological advances (Gouseti, 2010; McPherson & Bacow, 2015).

Today, artificial intelligence in education (Pinkwart, 2016), (computer-based) simulations (Chernikova et al., 2020), and augmented/mixed/virtual reality (Coban et al., 2022) are examples of the current progress in educational technologies (e.g. with the latest advances in prompt engineering for educational purposes; Mayer et al., 2022). On the one hand, it is assumed that new technologies will disrupt the current way of teaching and learning (K. Zhang & Aslan, 2021). On the other hand, traditional textbooks are still important as a medium of knowledge transfer for teachers and learners (Huß & Dölle, 2021). The assertion that AI will revolutionize the way people learn is similar to past claims. Every decade has seen new technologies touted as transformative for education (Ketamo, 2018). However, research shows that there is still room for improvement in the use of digital tools in VET (e.g., Mayer, Gentner, et al., 2023). In

recent years, technology integration has evolved from an initial resource-oriented phase, where the solution was simply to increase the number of computers, tablets, etc., to a pedagogy-oriented phase, where the focus is on effective teaching strategies (Mao, 2017). Teacher educators need to be aware of the pedagogical benefits of digital tools in their teaching and learning contexts in order to increase motivation through practical, effective, and professionally successful examples demonstrated by experienced teachers (Amhag et al., 2019). The question arises as to how student teachers learn to perform the professional work of a teacher in a digitally infused education system (Starkey, 2020). These changing teaching and learning environments require new forms of professional development for teachers (Mao, 2017), as the development of professional knowledge for teachers is seen as crucial for the effective integration of technology into business education (Seufert et al., 2019).

2.2 Teachers' professional knowledge

Pedagogical Content Knowledge

Teachers' professional knowledge can be operationalized in various ways (Fernandez, 2014). Pedagogical knowledge (PK), content knowledge (CK) and pedagogical content knowledge (PCK) are the elements of a professional knowledge model proposed by Shulman (1986) that is frequently used to describe the knowledge of teachers. CK refers to what teachers need to know about what they are teaching, whereas PK refers to what they know about teaching in general (Cochran et al., 1993). According to Shulman (1986), CK pertains to the understanding of concepts, theories, and ideas, as well as knowledge of proofs, evidence, practices, and methods for developing this knowledge. PK encompasses the educational objectives, teaching and learning methods, classroom techniques, understanding of the target audience, and strategies for assessing students' knowledge (Fernandez, 2014; Shulman, 1986). Theoretical models of PCK can be distinguished into transformative and integrative models (Gess-Newsome, 1999).

In an integrative model, PCK refers to the intersection of different areas of knowledge. Thus, PCK is based on the way teachers intertwine PK and CK (see Figure 2-1). According to Penso, "Pedagogical content knowledge is a type of knowledge that is unique to teachers, and is based on the manner in which teachers relate their pedagogical knowledge (what they know about teaching) to their subject matter knowledge (what they know about what they teach)" (2002, p. 25). PCK is seen as a type of knowledge that teachers develop through experience (Cochran et al., 1993). It includes knowledge of the main topics of the subject and how these can be organized and presented to facilitate learning. It also includes knowledge of learners and learning

processes, subject-specific learning difficulties such as typical misconceptions, learners' backgrounds and how these affect teaching (Penso, 2002; Shulman, 1986). Prospective teachers with little or no teaching experience tend to focus on subject knowledge, while knowledge of subject-specific pedagogy, learning environments and students' backgrounds are parts that require more teaching experience to develop (Cochran et al., 1993).

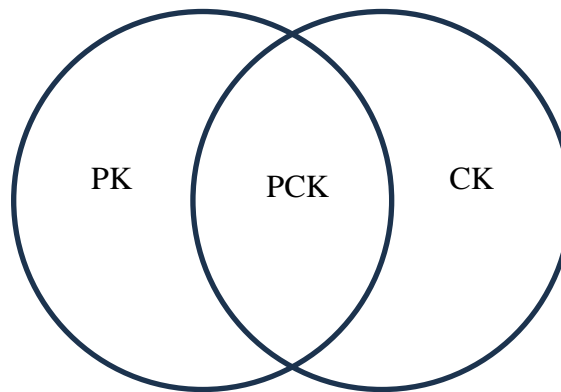


Figure 2-1: Integrative PCK model, adapted from Shulman (1986).

In a transformative model, PCK is “the transformation of subject matter knowledge, pedagogy and context knowledge to a distinct way - the only form of knowledge that would impact on teacher practice” (Fernandez, 2014, p. 94). Grossmann (1990) systematized teacher knowledge in a transformative model (see Figure 2-2) and distinguished four components: (1) subject matter knowledge, (2) general PK, (3) context knowledge, and (4) PCK as the outcome. First, subject matter knowledge refers to substantive structures (e.g., how to organize facts, fundamental concepts), content, and syntactic structures (e.g., how to create and promote knowledge). Second, general PK (Grossman, 1990) refers to knowledge about learners and learning, classroom management, curriculum and instruction, and others. Third, context knowledge is about students, their community, the district and school itself (Fernandez, 2014; Grossman, 1990). Finally, in this transformative model, PCK can be seen as the result of the transformation of teachers' professional knowledge. These traditional models are constrained by the particular characteristics of digital environments, emphasizing the need to integrate technology as part of teachers' professional knowledge.

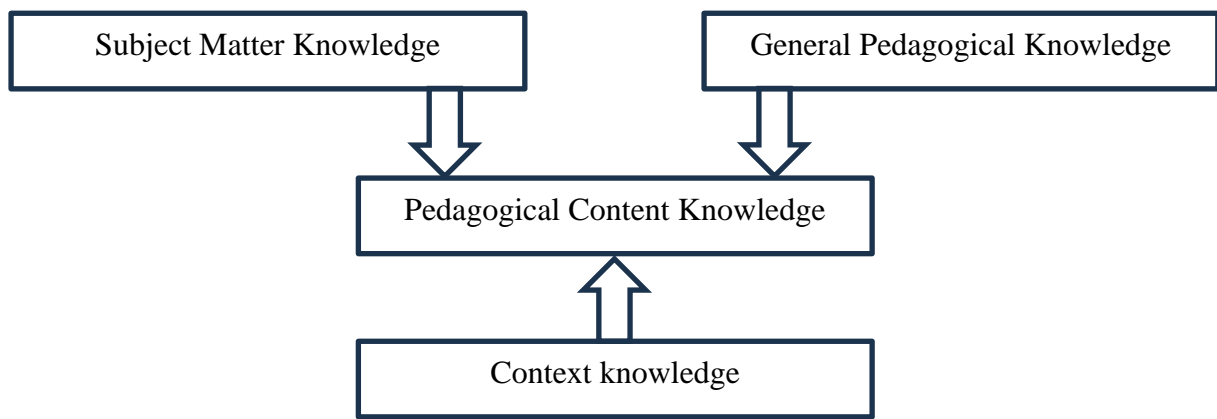


Figure 2-2: Transformative Model of teacher knowledge, adapted from Fernandez (2014; Grossmann 1990).

Technological Pedagogical Content Knowledge

In a digitalized educational system, technology can no longer be treated as a separate part of teachers' professional knowledge from the pedagogical and content knowledge that teachers need (W. Wang et al., 2018). The PCK model outlined so far lacks technological knowledge (TK), although digitalization is having a major impact on the education system and comparative studies such as the International Computer and Information Literacy Study (ICILS) have shown that TK needs to be promoted among pupils and teachers. (Bos et al., 2014). The TPACK model (see Figure 2-3; Koehler & Mishra, 2009; Mishra, 2019; Mishra & Koehler, 2006) therefore extends the previous model by TK. In the TPACK model, CK, PK and TK are the three fundamental dimensions that overlap as the basis dimensions should not be treated completely separated. Technology knowledge (TK) encompasses knowledge of all types of technology, from low-tech (e.g. chalkboards, worksheets) to high-tech (e.g. computers, virtual reality goggles). Technological content knowledge (TCK) enables teachers to select appropriate technologies for teaching and learning. TCK refers to knowing how technology changes the subject matter, for example, teaching with VR simulations to promote knowledge of the safety regulations in warehouse logistics. Technological pedagogical knowledge (TPK) is about how teaching methods can be adapted to the content and how technology can support these methods (e.g. simulations for business forecasting where students switch in a competitive scenario). Finally, TPACK, as the intersection of all previous knowledge areas, refers to knowledge about the flexible use of strategies and tools to effectively teach a particular content, e.g., by designing a curriculum that integrates ERP systems into business process education and provides students with both theoretical knowledge and hands-on experience through interactive simulations (e.g., Mayer & Seifried, 2024). In addition, the model emphasizes context as an important dimension of teachers' professional development (Mishra, 2019). This includes contextual support through

the provision of technological infrastructure, time and financial resources and didactic support through examples and guidance.

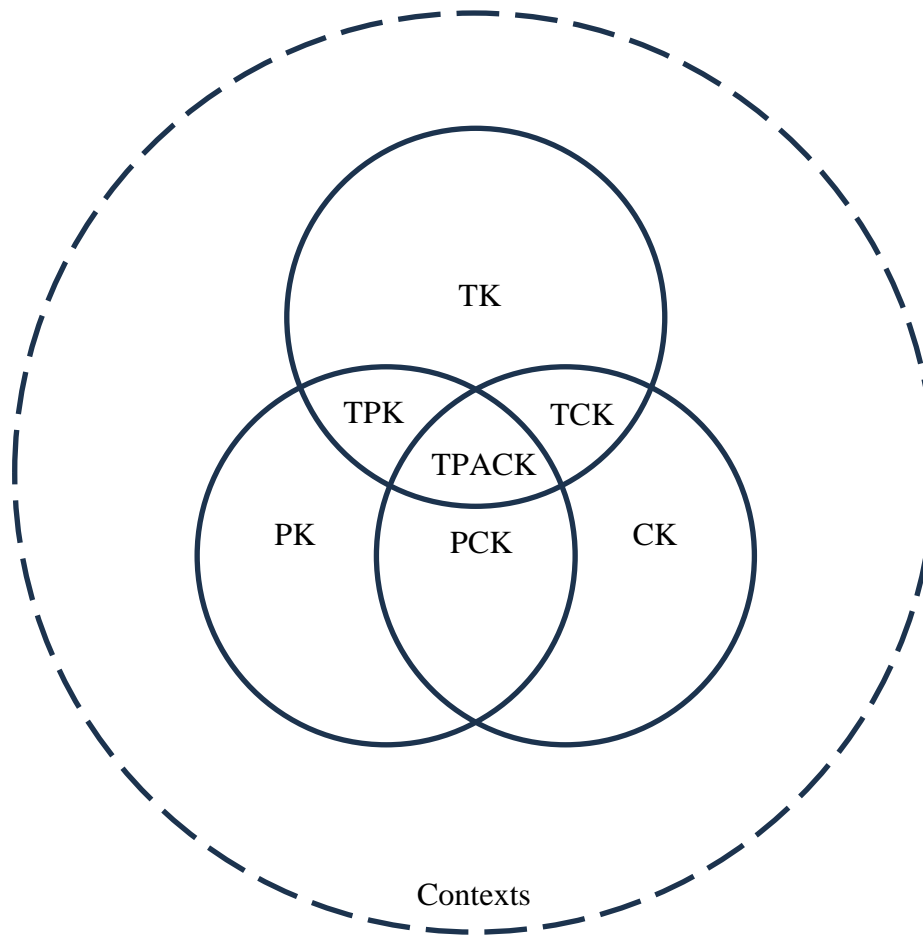


Figure 2-3: TPACK model, adapted from Mishra (2019), Mishra & Koehler (2006). Original Source: tpack.org.

Digitally-Related Pedagogical and Content Knowledge

The TPACK model is well established in the literature (Rosenberg & Koehler, 2015; Schmid et al., 2024; W. Zhang & Tang, 2021), but further models arise based on TPACK. In 2019, Huwer et al. introduced the digitality related pedagogical and content knowledge (DPACK) model, as an advancement of TPACK with a shift towards digital knowledge (DK). The authors applied the model on a domain-specific background for Science, Technology, Engineering and Mathematics (STEM) education (2019a, 2019b) and furthermore focuses on the analytical skills of teachers on the basis of professional knowledge. The DPACK model encompasses not only technical knowledge but also the ability to thoughtfully utilize and reflect the new possibilities offered by this technology (e.g., understanding new communication channels, utilizing digital sources). At around the same time, a similar but not equivalent DPACK model was presented by Döbeli Honegger (2021), which can be distinguished from the DPACK model of Huwer et

al (2019a). The model implements the Dagstuhl triangle (Brinda et al., 2016) as a critical, expanded view of a digitally networked world with a focus on related phenomena, objects and situations. This DPACK model reflects the impacts of digitization and further develops TPACK along the Dagstuhl triangle by taking into account a sociocultural perspective on digitalization (i.e., what are the impacts?), a technological perspective (i.e., how does it work?), and an application-based perspective (i.e., how do I use it?). Similarly, a paper by Thyssen et al. (2023) extends the DPACK model of Huwer et al. (2019a) by applying the Frankfurt triangle, as an extension of the Dagstuhl Triangle (Brinda et al., 2019). This extension of the Dagstuhl Triangle incorporates multiple disciplinary perspectives from researchers in media sociology, media theory, computer science and school practitioners. A key enhancement of DPACK the integration of a sociocultural knowledge (SC) sphere at the center of the model (see Figure 2-4). Sociocultural knowledge is considered an important part of teacher education and refers to intercultural communication skills, knowledge of social and cultural life, cultural understanding and language skills (Thyssen et al., 2023). This expansion creates numerous new overlaps¹ between sociocultural knowledge and the previously discussed TPACK areas. However, the recent DPACK models still require further operationalization and empirical validation.

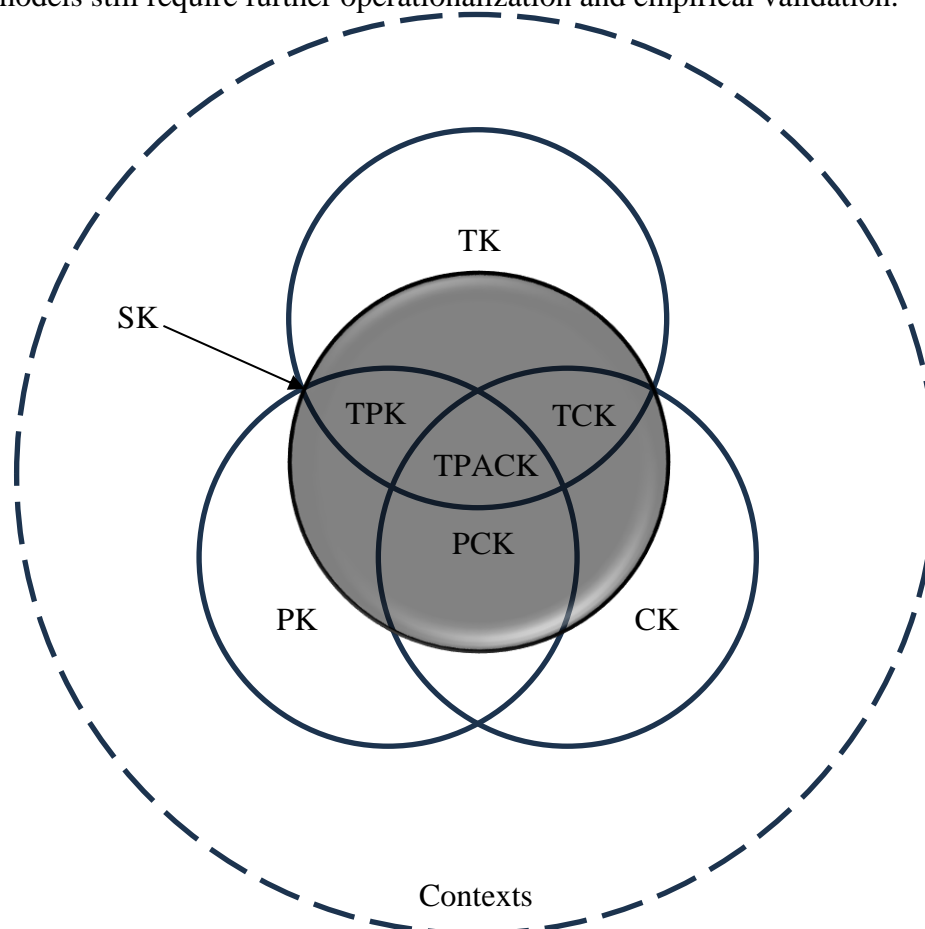


Figure 2-4: The DPACK model, adapted from Thyssen (2023) including a sociocultural knowledge sphere. The complexity has been reduced for illustrative purposes.

¹ For more details, see Thyssen et al. in their detailed discussion of the emerging overlaps (2023).

Therefore, this thesis adheres to the original TPACK model (Koehler & Mishra, 2009; Mishra, 2019; Mishra & Koehler, 2006), as one of the most commonly used models in educational research on teachers knowledge on teaching with technology. Current research also addresses several limitations that need to be overcome in TPACK research. For example, the unclear operationalization of knowledge and knowledge domains, a lack of experimental studies and longitudinal designs, as well as the lack of context-specific measures and the absence of student learning as an outcome variable, all are currently of interest to TPACK research (Schmid et al., 2024). Although TPACK research has continued to grow over the last 15 years, it is still unclear which of the conceptualizations of the relationships between the domains is more suitable, e.g. whether TPACK should be modeled in a transformative or an integrative way (Mouza et al., 2014; Schmid et al., 2020, 2024; Thyssen et al., 2023). This is crucial because the design of an effective TPACK intervention raises the question of whether the knowledge areas are interwoven or whether they should be treated independently of each other. In addition, the concepts and measures used for TPACK may differ between studies (e.g., assessment vs. self-report, neglect of related constructs such as self-efficacy, beliefs, competence). Furthermore, the impact of TPACK on learning outcomes (e.g., learner knowledge, student-teacher relationship) is often not examined and context-specific factors (e.g., occupational domain-background, teacher experience) are often not considered (Schmid et al., 2024). TPACK based training and related research is also frequently carried out in the field of VET (e.g., Sanger & Jenert, 2023; Schafer et al., 2020). Critics in the field of VET research point out that the dual content dimension of vocational education in Germany is not sufficiently taken into account and suggest a further development of TPACK that integrates vocational content (e.g. knowledge of daily business processes in different occupations) and disciplinary content (e.g. knowledge of business administration) as two parts of CK (Sanger & Jenert, 2023). However, the extensions of this model still need to be further operationalized and empirically tested.

Teachers' Digital Competencies

In addition to teacher professional knowledge models, digitalization as part of the teaching profession has also been integrated into various academic and policy frameworks that focus on teachers' digital competence. In general, Weinert (2001) defines the concept of competencies as a combination of “[...] intellectual abilities, content-specific knowledge, cognitive skills, domain-specific strategies, routines and subroutines, motivational tendencies, volitional control systems, personal value orientations, and social behaviours” (p. 51). According to Marín and Castañeda (2023), digital literacy is a crucial competence for educators, but the main challenge is the lack of a consistent terminology and a clear definition. Based on 33 reviews, the authors

conclude that digital literacy “emerges as a notion of situated multiple integrated skills and practices (conceptual, attitudinal, procedural, and ethical) that empower people (individuals and groups) to participate and communicate efficiently in society” (Marín & Castañeda, 2023, p. 1093). Two prominent examples are the United Nations Educational, Scientific and Cultural Organization's (UNESCO) ICT Competence Framework for Teachers (CFT) in its third version (UNESCO, 2018), and the European Digital Competence Framework for Teachers (DigCompEdu) from the European Commission's Joint Research Centre (Redecker, 2017).

The ICT CFT Version 3 aligns with the Agenda 2030 for Sustainable Development of the United Nations, updating relevant competencies to reflect current technological advances and evolving life and work demands. It includes open educational resources (OER) and emphasizes inclusive education, supporting the Sustainable Development Goals (SDG) principle of "leaving no one behind" (UNESCO, 2018, p. 7). Designed to inform policymakers and guide teacher-training policies and programs, it is applicable across all phases of teacher education, from pre-service to ongoing professional development. UNESCO experts have stated that teachers should be proficient in the use of new media in six areas at three levels of competence (knowledge acquisition, knowledge deepening, knowledge creation). The framework emphasizes the comprehensive use of ICT in the classroom (see Figure 2-5). The framework addresses: 1) understanding the changes in education; 2) designing curricula and assessing learning outcomes; 3) developing pedagogy with specific methodologies; 4) integrating digital literacy into teaching activities; 5) managing school facilities; and 6) facilitating professional development (UNESCO, 2018). Each area of the UNESCO framework includes objectives, required competences and practical skills. The term 'exemplary skills' is used intentionally to clarify that these vary according to educational level, teacher specialization and the changing information society (Tomczyk & Fedeli, 2021). The ICT Teacher Competence Framework includes an Open Educational Resource (OER) project to help develop teachers' digital competences. It aids UNESCO member states in adapting the framework and offers resources for teacher training via an OER hub curated by UNESCO.²

² See <https://www.oercommons.org/hubs/UNESCO>

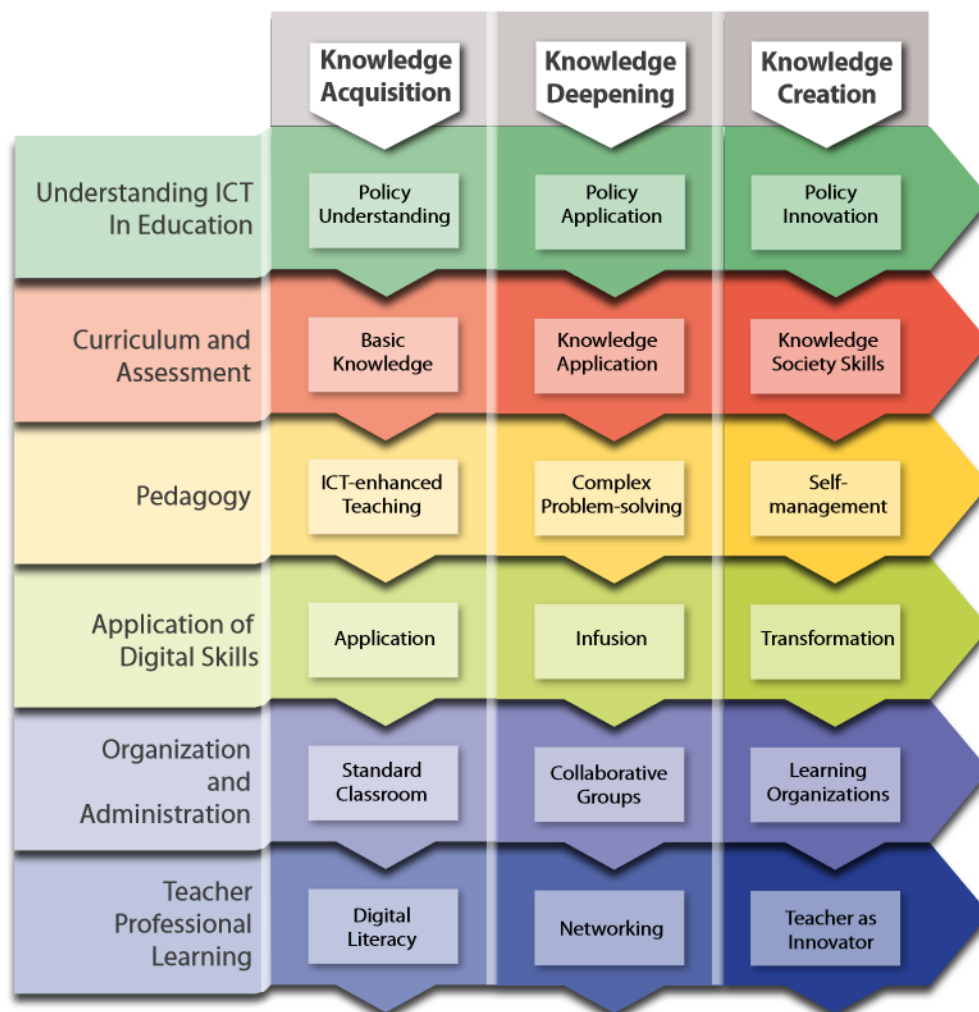


Figure 2-5: The UNESCO ICT CFT in its third version (UNESCO, 2018, p. 10).³

The European Framework DigCompEdu is a closely related framework to the UNESCO ICT CFT v3. DigCompEdu describes what it means to be digitally competent and supports the development of these competences across Europe. Here too, the aim is to inform both political decision-makers and practitioners. It applies to educators at all levels, from early childhood to adult education, including general and vocational education, special education and non-formal learning contexts (Redecker, 2017). It covers six key areas (see Figure 2-6): Professional Engagement, Digital Resources, Teaching and Learning, Assessment, Empowering Learners and Promoting Learners' Digital Competence. All areas contain specific indicators relating to learning and teaching activities, self-development and the wider context. It further categorizes digital literacy into six levels: Newcomer (A1), Explorer (A2), Integrator (B1), Expert (B2), Leader

³ Source: (UNESCO, 2018). UNESCO / UNESCO ICT Competency Framework for Teachers Version 3 / ISBN 978-92-3-100285-4 – licensed under CC BY SA. The present work (this thesis) is not an official UNESCO publication and shall not be considered as such.

(C1) and Pioneer (C2), each with detailed competence descriptions (Redecker, 2017). In addition, DigCompEdu offers a self-reflection tool named SELFIE, a questionnaire with 32 self-assessment questions that teachers can use to diagnose their own digital skills (Economou, 2023). The framework emphasizes that digital competence has an impact on the development of both learners' and teachers' (Tomczyk & Fedeli, 2021).

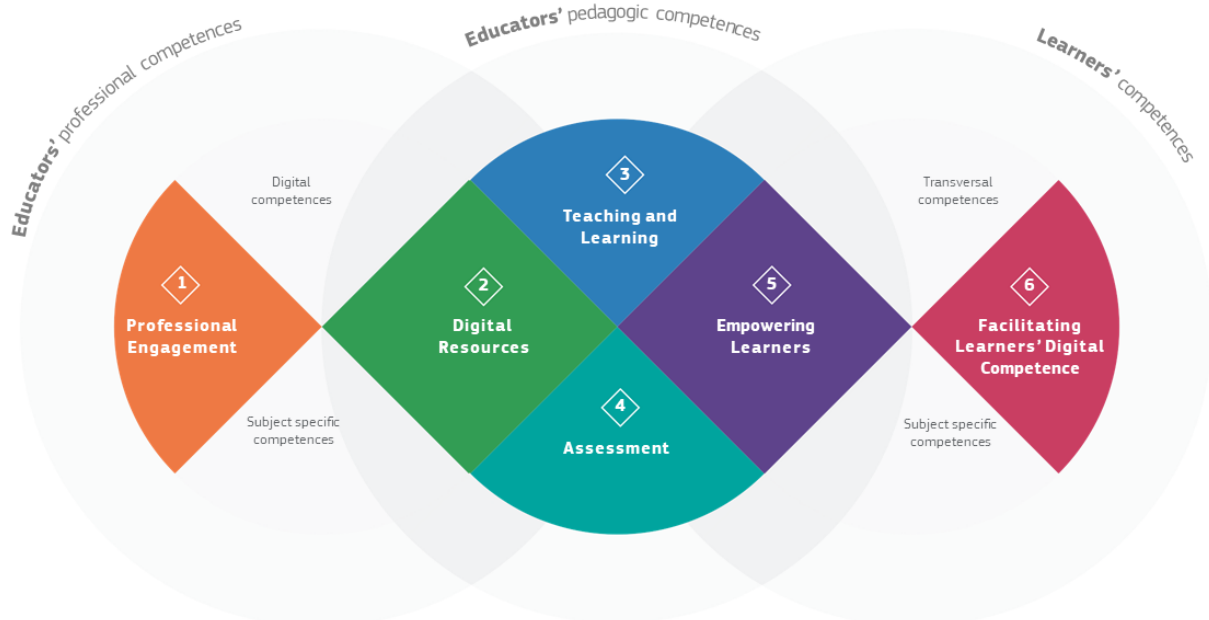


Figure 2-6: The DigCompEdu Framework Areas and Scope for the Digital Competence of Educators (Redecker, 2017, p. 15).⁴

Building on this, the following chapter explores how technological advances in research methods are providing deeper insights into learners' cognitive processes and engagement by analyzing their eye movements in technology-enhanced learning environments. This analysis can help refine instructional strategies and optimize the use of digital tools to further improve educational outcomes.

2.3 Analyzing learning processes in technological-rich environments

According to Euler and Wilbers (2020), digitalization not only enriches teaching and learning practice in VET, but also VET research from three perspectives: Firstly, technology-related

⁴ Source: First published, in English, in 2017, as "European Framework for the Digital Competence of Educators: DigCompEdu" by the European Commission's Joint Research Centre, EUR 28775 EN, ISBN 978-92-79-73494-6, doi:10.2760/159770, JRC107466, <http://europa.eu/!gt63ch>.

The European Commission is not responsible for the present work (this thesis) and cannot be held liable for any consequence stemming from its use.

research should be supported by trend monitoring, which serves as a descriptive record of existing developments at the various levels of vocational education and training. Secondly, research should be conducted into the consequences of the use of digital technologies, with a view to establishing normative, economic, and social criteria for their use. Thirdly, the actual use of digital tools and their empirically proven impact on learning should be explored. This thesis focuses on the third perspective and emphasizes empirical research in technology-intensive environments where modern learning technologies are used.

Although the term ‘educational technology’ is typically understood to encompass a range of media and devices, many scholars and practitioners tend to view the discipline more narrowly as the utilization of information and communication technology in educational settings (Niegemann & Weinberger, 2020). Advanced learning technologies enable, support, or enhance learning (Cerri, 2012). The term ‘advanced’ can be defined as “dynamic, experimental, to be implemented and evaluated in order to limit the risk that what we describe today as advanced will be considered obsolete in a few months” (Cerri, 2012, p. 154). This thesis presents advanced educational technologies, including interactive learning environments and computer-based simulations. An interactive learning environment is defined as a software (or, in some cases, hardware) that facilitates teaching and learning through interaction between learners and the system, between teachers and the system, or between teachers and learners using the system. This technology can be used to facilitate academic, informal, or work-based learning, and it can range from passive virtual worlds to active intelligent tutoring systems (Psocka, 2012). Computer-based simulations refer to simulation learning, i.e., “a form of experiential learning that is learner-centred, integrates many facets of learning (e.g. cognitive, motivational, affective, psychomotor, social) and has a high degree of authenticity” (Breckwoldt et al., 2014, p. 674). Computer simulations provide a secure learning environment for learning (i.e., a safe space for training purposes) by implementing complex environments that are highly realistic and mirror reality through dynamic interactions, changes over time, and the interplay of variables in domain-specific circumstances. This allows for the simulation of the often complex decision-making and problem-solving processes involved in professional tasks (Breckwoldt et al., 2014; Dörner & Funke, 2017). VET research has so far focused primarily on outcomes such as skills assessment and learning performance rather than learning processes (Abele, 2018), which was often due to a lack of non-intrusive and easy-to-implement technical options.

Technological advances can foster the scientific understanding of learning processes in interactive learning environments and computer-based simulations. Research on learning processes

in technology-rich environments can provide a more profound comprehension of the way teachers and learners interact with technology, the effective implementation of technology-based instructional strategies, and the enhancement of learning content areas through the utilization of domain-specific digital tools. Eye-tracking represents one such research method for the collection of process data in the form of eye movements (Duchowski, 2017; Holmqvist et al., 2011; Holmqvist & Andersson, 2017). The method has evolved into a cost-effective and non-intrusive online (i.e., process) measurement of learning behavior. Eye-tracking data provides information about eye movements, including blinks, pupil dilation, fixations, and saccades. Fixations occur when the eyes remain more or less motionless⁵ on a particular object, while saccades are the rapid movements of the eyes between fixations.

Eye-tracking has become increasingly important in educational research as it has improved computer-based learning, helped to understand how expertise develops in visual domains (e.g., chess), and is being used to promote visual expertise of novices through eye-movement records of experts (Jarodzka et al., 2017). The method provides behavioral data on visual attention and learning processes that shed light on how students interact with digital learning materials to inform and improve pedagogical practice. Eye movement measurements can be used to assess the cognitive processes involved in learning with multimedia (Alemdag & Cagiltay, 2018) and a particular focus can be placed on domain-specific problem solving in computer-based simulations for VET (e.g., Mayer, Rausch, et al., 2023). For example, the search for information may be indicated by the initial fixation time on a specific area of interest (AOI), while the processing of information can be linked to the average fixation duration on an AOI. In addition, the frequency of transitions between different AOIs in advanced learning technologies can provide insights into the process of working out how to solve a particular problem. However, the interpretation of eye-tracking data is not always straightforward. For example, it is unclear whether a long fixation duration indicates deep cognitive processing and elaboration or, conversely, confusion or behavior that indicates abandonment or superficial processing. As many readers have probably noticed, reading can occur with varying degrees of processing. Eye-tracking research must therefore be based on a theoretical foundation and supplemented and triangulated by additional online and offline measurements such as questionnaires or (retrospective) think-aloud.

Although, eye tracking is more of a research method that is not directly related to teachers' professional knowledge, the results can have a significant impact on the teaching profession,

⁵ In fact, the eye does not remain completely still, but fixates through micromovements: slow drift movements, fast microsaccades, and tremor (Møller et al., 2006).

e.g., with regard to formative assessment (Y. Wang et al., 2021), professional vision (Keskin et al., 2024), or by analyzing instructional design (Alemdag & Cagiltay, 2018). Eye-tracking data can assess students' visual attention and could link online measurements to understanding learning content in technology-rich environments (e.g., Mayer et al., under review). Teachers can be supported with this information to adapt lessons and provide targeted support to struggling learners (e.g., Buettner et al., 2018). A profound branch of eye-tracking research is related to the professional vision of prospective and experienced teachers (e.g., Grub et al., 2022; Huang et al., 2023; Wyss et al., 2020). The implementation of interventions using eye-tracking recordings by experienced teachers or students' own recordings could promote professional vision (e.g., Telgmann & Müller, 2023). Finally, if teachers understand how students visually engage with learning materials, they can design more effective teaching materials. For example, if eye-tracking shows that students struggle with certain visualizations, teachers can modify these materials to improve clarity and engagement (e.g., Coskun & Cagiltay, 2022).

Eye-tracking technology, although at first glance only marginally related to the teaching profession, provides valuable insights that can significantly improve the profession and therefore fits within the conceptual framework of this thesis. However, eye-tracking research also creates new opportunities for insights, but also new responsibilities for researchers with various challenges, such as methodological differences between studies (e.g. domain-specific content and different teaching styles and study participants) or the ethical aspects of data recording (Jarodzka et al., 2017). Furthermore, research with learning environments based on current technological advances often runs the risk of becoming outdated due to the rapid innovation of technologies (Euler & Wilbers, 2020).

2.4 Summary

As described in this chapter, the process of digitalization, when considered as a socio-technical phenomenon, has a meaningful impact on the field of VET. As evidenced by a brief historical overview, the efficacy of educational technologies has often been overstated. The ineffectiveness observed in practice can be attributed at least in part to a lack of knowledge and training in this area. There is a discrepancy between pedagogical theory and practice, and research should address the challenge of making pedagogical theory relevant to student teachers (Puustinen et al., 2018). In this context, TPACK, a model for prospective teachers' professional knowledge, can help to elucidate the deficiencies in knowledge that must be addressed to facilitate digitalization in VET. Furthermore, technological advancement is also transforming VET research. Given that this work is partly based on learning process research in advanced digital learning environments, the methodology of eye tracking, as a specific online measurement of

learning behavior, may prove valuable in elucidating learning processes in VET and informing teaching and learning in practice in light of the increasing digitization of education.

2.5 References

- Abele, S. (2018). Diagnostic Problem-Solving Process in Professional Contexts: Theory and Empirical Investigation in the Context of Car Mechatronics Using Computer-Generated Log-Files. *Vocations and Learning*, *11*(1), 133–159. <https://doi.org/10.1007/s12186-017-9183-x>
- Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, *125*, 413–428. <https://doi.org/10.1016/j.compedu.2018.06.023>
- Amhag, L., Hellström, L., & Stigmar, M. (2019). Teacher Educators' Use of Digital Tools and Needs for Digital Competence in Higher Education. *Journal of Digital Learning in Teacher Education*, *35*(4), 203–220. <https://doi.org/10.1080/21532974.2019.1646169>
- Bennett, S., Maton, K., & Kervin, L. (2008). The 'digital natives' debate: A critical review of the evidence. *British Journal of Educational Technology*, *39*(5), 775–786. <https://doi.org/10.1111/j.1467-8535.2007.00793.x>
- Bos, W., Eickelmann, B., Gerick, J., Goldhammer, F., Schaumburg, H., Schwippert, K., Senkbeil, M., Schulz-Zander, R., & Wendt, H. (2014). *ICILS 2013: Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern in der 8. Jahrgangsstufe im internationalen Vergleich [ICILS 2013: International comparison of computer and information-related skills of pupils in year 8]*. Waxmann.
- Breckwoldt, J., Gruber, H., & Wittmann, A. (2014). Simulation Learning. In S. Billett, C. Harteis, & H. Gruber (Eds.), *International Handbook of Research in Professional and Practice-based Learning* (pp. 673–698). Springer. https://doi.org/10.1007/978-94-017-8902-8_25
- Brinda, T., Brügggen, N., Diethelm, I., Knaus, T., Kommer, S., Kopf, C., Missomelius, P., Leschke, R., Tilemann, F., & Weich, A. (2019). Frankfurt-Dreieck zur Bildung in der digital vernetzten Welt: Ein interdisziplinäres Modell [Frankfurt triangle for education in the digitally networked world: an interdisciplinary model]. *18. GI-Fachtagung Informatik und Schule - Informatik für alle*, 25–33. <https://doi.org/10.18420/INFOS2019-A1>
- Brinda, T., Diethelm, I., Gemulla, R., Romeike, R., Schöning, J., Schulte, C., Bartoschek, T., Bergner, N., Dietrich, L., Döbeli, B., Fries, R., Hellmig, L., Herzig, B., Hollatz, J., Jörisen, B., Kommer, S., Mittag, A., Kusterer, P., Oberweis, A., ... Zimmol, M. (2016). *Dagstuhl-Erklärung: Bildung in der digitalen vernetzten Welt*. Gesellschaft für Informatik e.V. <https://dl.gi.de/handle/20.500.12116/44026>
- Buettner, R., Sauer, S., Maier, C., & Eckhardt, A. (2018). Real-time Prediction of User Performance based on Pupillary Assessment via Eye-Tracking. *AIS Transactions on Human-Computer Interaction*, *10*(1), 26–60. <https://doi.org/10.17705/1thci.00103>
- Cerri, S. A. (2012). Advanced Learning Technologies. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 154–157). Springer. https://doi.org/10.1007/978-1-4419-1428-6_57
- Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-Based Learning in Higher Education: A Meta-Analysis. *Review of Educational Research*, *90*(4), 499–541. <https://doi.org/10.3102/0034654320933544>

- Coban, M., Bolat, Y. I., & Goksu, I. (2022). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review*, 36, 100452. <https://doi.org/10.1016/j.edurev.2022.100452>
- Cochran, K. F., DeRuiter, J. A., & King, R. A. (1993). Pedagogical Content Knowing: An Integrative Model for Teacher Preparation. *Journal of Teacher Education*, 44(4), 263–272. <https://doi.org/10.1177/0022487193044004004>
- Coskun, A., & Cagiltay, K. (2022). A systematic review of eye-tracking-based research on animated multimedia learning. *Journal of Computer Assisted Learning*, 38(2), 581–598. <https://doi.org/10.1111/jcal.12629>
- Döbeli Honegger, B. (2021). Covid-19 und die digitale Transformation in der Schweizer Lehrerinnen- und Lehrerbildung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 39(3), 411–422. <https://doi.org/10.25656/01:23693>
- Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, 8, 1153. <https://doi.org/10.3389/fpsyg.2017.01153>
- Duchowski, A. T. (2017). *Eye Tracking Methodology*. Springer. <https://doi.org/10.1007/978-3-319-57883-5>
- Economou, A. (2023). *SELFIEforTEACHERS. Designing and developing a self-reflection tool for teachers' digital competence*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2760/40528>
- Euler, D., & Wilbers, K. (2020). Berufsbildung in digitalen Lernumgebungen [Vocational training in digital learning environments]. In R. Arnold, A. Lipsmeier, & M. Rohs (Eds.), *Handbuch Berufsbildung* (pp. 427–438). Springer. https://doi.org/10.1007/978-3-658-19312-6_34
- Evans, C., & Robertson, W. (2020). The four phases of the digital natives debate. *Human Behavior and Emerging Technologies*, 2(3), 269–277. <https://doi.org/10.1002/hbe2.196>
- Fernandez, C. (2014). Knowledge Base for Teaching and Pedagogical Content Knowledge (PCK): Some Useful Models and Implications for Teachers' Training. *Problems of Education in the 21st Century*, 60(1), 79–100. <https://doi.org/10.33225/pec/14.60.79>
- Frenzel-Piasentin, A., Muench, J., Bruckner, M., & Veit, D. (2021). Digitization or Digitalization? – Toward an Understanding of Definitions, Use and Application in IS Research. *AMCIS 2021 Proceedings*, 18. https://aisel.aisnet.org/amcis2021/adv_info_systems_general_track/adv_info_systems_general_track/18
- Gal, P., Nicoletti, G., Renault, T., Sorbe, S., & Timiliotis, C. (2019). *Digitalisation and productivity: In search of the holy grail – Firm-level empirical evidence from EU countries* (OECD Economics Department Working Papers 1533). OECD Publishing. <https://doi.org/10.1787/5080f4b6-en>
- Gess-Newsome, J. (1999). Pedagogical Content Knowledge: An Introduction and Orientation. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining Pedagogical Content Knowledge* (pp. 3–17). Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47217-1_1
- Gorshid, G. D., Mayer, C., Rausch, A., & Seifried, J. (2022). Das LUCA-Dashboard im Usability-Test-Eine gaze-cued retrospective Think-Aloud-Studie [The LUCA dashboard in a usability test—A retrospective think-aloud study]. In S. Schumann, S. Seeber, & S. Abele (Eds.), *Digitale Transformation in der Berufsbildung* (pp. 189–212). wbv.

- Gouseti, A. (2010). Web 2.0 and education: Not just another case of hype, hope and disappointment? *Learning, Media and Technology*, 35(3), 351–356. <https://doi.org/10.1080/17439884.2010.509353>
- Gradillas, M., & Thomas, L. D. W. (2023). Distinguishing digitization and digitalization: A systematic review and conceptual framework. *Journal of Product Innovation Management, Early View*, 1–32. <https://doi.org/10.1111/jpim.12690>
- Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. Teachers College Press.
- Grub, A.-S., Biermann, A., Lewalter, D., & Brünken, R. (2022). Professional Vision and the Compensatory Effect of a Minimal Instructional Intervention: A Quasi-Experimental Eye-Tracking Study With Novice and Expert Teachers. *Frontiers in Education*, 7, 890690. <https://doi.org/10.3389/educ.2022.890690>
- Helsper, E. J., & Eynon, R. (2010). Digital natives: Where is the evidence? *British Educational Research Journal*, 36(3), 503–520. <https://doi.org/10.1080/01411920902989227>
- Holmqvist, K., & Andersson, R. (2017). *Eye tracking: A comprehensive guide to methods, paradigms, and measures* (2nd edition). CreateSpace.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- Huang, Y., Miller, K. F., Cortina, K. S., & Richter, D. (2023). Teachers' professional vision in action: Comparing expert and novice teacher's real-life eye movements in the classroom. *Zeitschrift Für Pädagogische Psychologie*, 37(1–2), 122–139. <https://doi.org/10.1024/1010-0652/a000313>
- Huß, B., & Dölle, F. (2021). *Bedeutung, Nutzung und Zugang zu Lehrbüchern an Hochschulen: Zentrale Ergebnisse der Befragungsstudie für das Bundesministerium für Bildung und Forschung (BMBF): Bericht Dezember 2021 [Importance, use and access to textbooks at universities : key findings of the survey study for the Federal Ministry of Education and Research (BMBF): Report December 2021]*. Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). <https://doi.org/10.2314/KXP:1788361857>
- Huwer, J., Irion, T., Kuntze, S., Schaal, S., & Thyssen, C. (2019a). From TPaCK to DPaCK—Digitalization in education requires more than technical knowledge. In M. Shelley & S. A. Kiray (Eds.), *Education research highlights in mathematics, science and technology* (pp. 298–309). ISRES Publishing.
- Huwer, J., Irion, T., Kuntze, S., Schaal, S., & Thyssen, C. (2019b). Von TPaCK zu DPaCK—Digitalisierung im Unterricht erfordert mehr als technisches Wissen [From TPaCK to DPaCK digitization in the classroom requires more than technical knowledge]. *MNU Journal*, 72(5), 358–364.
- Hwang, G.-J., & Chang, C.-Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 31(7), 4099–4112. <https://doi.org/10.1080/10494820.2021.1952615>
- Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in Educational Science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research*, 10(1), 1–18. <https://doi.org/10.16910/JEMR.10.1.3>
- Keskin, Ö., Seidel, T., Stürmer, K., & Gegenfurtner, A. (2024). Eye-tracking research on teacher professional vision: A meta-analytic review. *Educational Research Review*, 42, 100586. <https://doi.org/10.1016/j.edurev.2023.100586>

- Ketamo, H. (2018). *Dreams and reality: How AI will change education*. UNESCO: United Nations Educational, Scientific and Cultural Organisation. <https://mgiep.unesco.org/article/dreams-and-reality-how-ai-will-change-education>
- Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60–70.
- Legner, C., Eymann, T., Hess, T., Matt, C., Böhmman, T., Drews, P., Mädche, A., Urbach, N., & Ahlemann, F. (2017). Digitalization: Opportunity and Challenge for the Business and Information Systems Engineering Community. *Business & Information Systems Engineering*, 59(4), 301–308. <https://doi.org/10.1007/s12599-017-0484-2>
- Leiß, T. V., Rausch, A., & Seifried, J. (2022). Problem-Solving and Tool Use in Office Work: The Potential of Electronic Performance Support Systems to Promote Employee Performance and Learning. *Frontiers in Psychology*, 13, 869428. <https://doi.org/10.3389/fpsyg.2022.869428>
- Mao, J. (2017). New Technologies and Old Professional Development: A Reflection on Emerging Approaches. In W. W. K. Ma, C.-K. Chan, K. Tong, H. Fung, & C. W. R. Fong (Eds.), *New Ecology for Education—Communication X Learning* (pp. 67–77). Springer. https://doi.org/10.1007/978-981-10-4346-8_6
- Marín, V. I., & Castañeda, L. (2023). Developing Digital Literacy for Teaching and Learning. In O. Zawacki-Richter & I. Jung (Eds.), *Handbook of Open, Distance and Digital Education* (pp. 1089–1108). Springer. https://doi.org/10.1007/978-981-19-2080-6_64
- Mayer, C. W., Findeisen, S., Guggemos, J., & Seifried, J. (under review). Visual attention while solving economic learning tasks using dynamic graphs – An eye-tracking study. *Journal of Economic Education*.
- Mayer, C. W., Gentner, S., & Seifried, J. (2023). Digitale Unterrichtspraxis an kaufmännischen Schulen in der Corona-Pandemie—Eine Momentaufnahme [Digital teaching practice at commercial schools during the corona pandemic - a snapshot]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 119(3), 531–552. <https://doi.org/10.25162/zbw-2023-0021>
- Mayer, C. W., Ludwig, S., & Brandt, S. (2022). Prompt text classifications with transformer models! An exemplary introduction to prompt-based learning with large language models. *Journal of Research on Technology in Education*, 55(1), 1–17. <https://doi.org/10.1080/15391523.2022.2142872>
- Mayer, C. W., Rausch, A., & Seifried, J. (2023). Analysing domain-specific problem-solving processes within authentic computer-based learning and training environments by using eye-tracking: A scoping review. *Empirical Research in Vocational Education and Training*, 15(1), 2. <https://doi.org/10.1186/s40461-023-00140-2>
- Mayer, C. W., & Seifried, J. (2022). Lehren und Lernen mit ERP-Systemen – Befunde der Evaluation einer Lehrveranstaltung zur Förderung der Technologiekompetenz von Studierenden der Wirtschaftspädagogik [Teaching and learning with ERP systems—Findings from the evaluation of a course to promote the technological competence of business education students]. In U. Mathis, N. Ondrusch, D. Kilian, H. Krcmar, K. Turowski, & S. Weidner (Eds.), *Proceedings of the SAP Academic Community Conference 2022 DACH* (pp. 10–24). Technische Universität München. <https://mediatum.ub.tum.de/doc/1685828/document.pdf#page=10>
- Mayer, C. W., & Seifried, J. (2024). Looking behind the scenes—Being able to master ERP systems as a goal of vocational education and training. In Cedefop & OECD (Eds.), *Apprenticeships and the digital transition* (pp. 118–127). Publications Office.

- McPherson, M. S., & Bacow, L. S. (2015). Online Higher Education: Beyond the Hype Cycle. *Journal of Economic Perspectives*, 29(4), 135–154. <https://doi.org/10.1257/jep.29.4.135>
- Mishra, P. (2019). Considering Contextual Knowledge: The TPACK Diagram Gets an Upgrade. *Journal of Digital Learning in Teacher Education*, 35(2), 76–78. <https://doi.org/10.1080/21532974.2019.1588611>
- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record: The Voice of Scholarship in Education*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Molenda, M. H. (2022). History and Development of Instructional Design and Technology. In O. Zawacki-Richter & J. Insung (Eds.), *Handbook of Open, Distance and Digital Education* (pp. 1–18). Springer. https://doi.org/10.1007/978-981-19-0351-9_4-1
- Møller, F., Laursen, M. L., & Sjølie, A. K. (2006). The contribution of microsaccades and drifts in the maintenance of binocular steady fixation. *Graefes Archive for Clinical and Experimental Ophthalmology*, 244(4), 465–471. <https://doi.org/10.1007/s00417-004-0989-5>
- Mouza, C., Karchmer-Klein, R., Nandakumar, R., Yilmaz Ozden, S., & Hu, L. (2014). Investigating the impact of an integrated approach to the development of preservice teachers' technological pedagogical content knowledge (TPACK). *Computers & Education*, 71, 206–221. <https://doi.org/10.1016/j.compedu.2013.09.020>
- Mugge, P., Abbu, H., Michaelis, T. L., Kwiatkowski, A., & Gudergan, G. (2020). Patterns of Digitization: A Practical Guide to Digital Transformation. *Research-Technology Management*, 63(2), 27–35. <https://doi.org/10.1080/08956308.2020.1707003>
- Niegemann, H. (2020). Introduction—A Look Back Ahead. In E. Wuttke, J. Seifried, & H. Niegemann (Eds.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (pp. 15–19). Barbara Budrich. <https://doi.org/10.2307/j.ctv18dvv1c>
- Niegemann, H., & Weinberger, A. (2020). Was ist Bildungstechnologie? [What is educational technology?]. In H. Niegemann & A. Weinberger (Eds.), *Handbuch Bildungstechnologie* (pp. 3–16). Springer. https://doi.org/10.1007/978-3-662-54368-9_1
- Parviainen, P., Tihinen, M., Kääriäinen, J., & Teppola, S. (2017). Tackling the digitalization challenge: How to benefit from digitalization in practice. *International Journal of Information Systems and Project Management*, 5(1), 63–77. <https://doi.org/10.12821/ijispm050104>
- Penso, S. (2002). Pedagogical Content Knowledge: How do student teachers identify and describe the causes of their pupils' learning difficulties? *Asia-Pacific Journal of Teacher Education*, 30(1), 25–37. <https://doi.org/10.1080/13598660120114959>
- Pinkwart, N. (2016). Another 25 Years of AIED? Challenges and Opportunities for Intelligent Educational Technologies of the Future. *International Journal of Artificial Intelligence in Education*, 26(2), 771–783. <https://doi.org/10.1007/s40593-016-0099-7>
- Prensky, M. (2001a). Digital Natives, Digital Immigrants Part 1. *On the Horizon*, 9(5), 1–6. <https://doi.org/10.1108/10748120110424816>
- Prensky, M. (2001b). Digital Natives, Digital Immigrants Part 2: Do They Really Think Differently? *On the Horizon*, 9(6), 1–6. <https://doi.org/10.1108/10748120110424843>

- Psocka, J. (2012). Interactive Learning Environments. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 1604–1606). Springer. https://doi.org/10.1007/978-1-4419-1428-6_321
- Puustinen, M., Säntti, J., Koski, A., & Tammi, T. (2018). Teaching: A practical or research-based profession? Teacher candidates' approaches to research-based teacher education. *Teaching and Teacher Education*, *74*, 170–179. <https://doi.org/10.1016/j.tate.2018.05.004>
- Rausch, A., Deutscher, V., Seifried, J., Brandt, S., & Winther, E. (2021). Die web-basierte Bürosimulation LUCA – Funktionen, Einsatzmöglichkeiten und Forschungsausblick [The web-based office simulation LUCA - functions, possible applications and research outlook]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *117*(3), 372–394. <https://doi.org/10.25162/zbw-2021-0017>
- Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu*. (Y. Punie, Ed.). Publications Office of the European Union. <https://data.europa.eu/doi/10.2760/178382>
- Reich, J. (2020). *Failure to disrupt: Why technology alone can't transform education*. Harvard University Press.
- Rosenberg, J. M., & Koehler, M. J. (2015). Context and Technological Pedagogical Content Knowledge (TPACK): A Systematic Review. *Journal of Research on Technology in Education*, *47*(3), 186–210. <https://doi.org/10.1080/15391523.2015.1052663>
- Sänger, N., & Jenert, T. (2023). TPACK, DPACK, XY-PACK?: Eine kritische Auseinandersetzung mit der Modellierung digitaler Inhaltskompetenzen für die berufliche Bildung [TPACK, DPACK, XY-PACK?: A Critical Examination of the Modeling of Digital Content Competencies for Vocational Education]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *119*(3), 434–454. <https://doi.org/10.25162/zbw-2023-0017>
- Schäfer, P., Link, N., & Walker, F. (2020). Assessing Professional Knowledge of Teachers at Vocational Schools—Using the Example of a Professional Development for Automation and Digitized Production. In E. Wuttke, J. Seifried, & H. Niegemann (Eds.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (pp. 131–154). Barbara Budrich. <https://doi.org/10.2307/j.ctv18dvv1c>
- Schmid, M., Brianza, E., Mok, S. Y., & Petko, D. (2024). Running in circles: A systematic review of reviews on technological pedagogical content knowledge (TPACK). *Computers & Education*, *214*, 105024. <https://doi.org/10.1016/j.compedu.2024.105024>
- Schmid, M., Brianza, E., & Petko, D. (2020). Developing a short assessment instrument for Technological Pedagogical Content Knowledge (TPACK.xs) and comparing the factor structure of an integrative and a transformative model. *Computers & Education*, *157*, 103967. <https://doi.org/10.1016/j.compedu.2020.103967>
- Seufert, S., Guggemos, J., Tarantini, E., & Schumann, S. (2019). Professionelle Kompetenzen von Lehrpersonen im Kontext des digitalen Wandels: Entwicklung eines Rahmenkonzepts und Validierung in der kaufmännischen Domäne [Professional competencies of teachers in the context of digital change: development of a framework concept and validation in the commercial domain]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *115*(2), 312–339. <https://doi.org/10.25162/zbw-2019-0013>
- Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, *15*(2), 4–14. <https://doi.org/10.3102/0013189X015002004>

- Smith, E. E., Kahlke, R., & Judd, T. (2020). Not just digital natives: Integrating technologies in professional education contexts. *Australasian Journal of Educational Technology*, 36(3), 1–14. <https://doi.org/10.14742/ajet.5689>
- Smith, F. J. (1913). The Evolution of the Motion Picture: VI—Looking into the Future with Thomas A. Edison. In *The New York Dramatic Mirror* (pp. 24, Column 3).
- Spener, C., & Schumann, S. (2021). Wissensseffekte des ERP-Einsatzes in der kaufmännischen Berufsschule [Knowledge effects of ERP use in commercial vocational schools]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117(3), 395–430. <https://doi.org/10.25162/zbw-2021-0018>
- Starkey, L. (2020). A review of research exploring teacher preparation for the digital age. *Cambridge Journal of Education*, 50(1), 37–56. <https://doi.org/10.1080/0305764X.2019.1625867>
- Telgmann, L., & Müller, K. (2023). Training & prompting pre-service teachers' noticing in a standardized classroom simulation – a mobile eye-tracking study. *Frontiers in Education*, 8, 1266800. <https://doi.org/10.3389/educ.2023.1266800>
- Thyssen, C., Huwer, J., Irion, T., & Schaal, S. (2023). From TPACK to DPACK: The “Digitality-Related Pedagogical and Content Knowledge”-Model in STEM-Education. *Education Sciences*, 13(8), 769. <https://doi.org/10.3390/educsci13080769>
- Tomczyk, L., & Fedeli, L. (2021). Digital Literacy among Teachers—Mapping Theoretical Frameworks: TPACK, DigCompEdu, UNESCO, NETS-T, DigiLit Leicester. *Proceedings of the 38th International Business Information Management Association (IBIMA)*, 244–252.
- UNESCO. (2018). *UNESCO ICT Competency Framework for Teachers: Version 3*. UNESCO.
- Urlaub, P., & Dessen, E. (2022). Machine translation and foreign language education. *Frontiers in Artificial Intelligence*, 5, 936111. <https://doi.org/10.3389/frai.2022.936111>
- Vrana, J., & Singh, R. (2021). Digitization, Digitalization, and Digital Transformation. In N. Meyendorf, N. Ida, R. Singh, & J. Vrana (Eds.), *Handbook of Nondestructive Evaluation 4.0* (pp. 1–17). Springer. https://doi.org/10.1007/978-3-030-48200-8_39-1
- Wang, W., Schmidt-Crawford, D., & Jin, Y. (2018). Preservice Teachers' TPACK Development: A Review of Literature. *Journal of Digital Learning in Teacher Education*, 34(4), 234–258. <https://doi.org/10.1080/21532974.2018.1498039>
- Wang, Y., Lu, S., & Harter, D. (2021). Towards Collaborative and Intelligent Learning Environments Based on Eye Tracking Data and Learning Analytics: A Survey. *IEEE Access*, 9, 137991–138002. <https://doi.org/10.1109/ACCESS.2021.3117780>
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Hogrefe.
- Wuttke, E., Seifried, J., & Niegemann, H. (2020). Editorial: Vocational Education and Training in the Age of Digitization—Challenges and Opportunities. In E. Wuttke, J. Seifried, & H. Niegemann (Eds.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (pp. 15–19). Barbara Budrich. <https://doi.org/10.2307/j.ctv18dvv1c>
- Wyss, C., Rosenberger, K., & Bühner, W. (2020). Student Teachers' and Teacher Educators' Professional Vision: Findings from an Eye Tracking Study. *Educational Psychology Review*, 33, 91–107. <https://doi.org/10.1007/s10648-020-09535-z>

- Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 2, 100025. <https://doi.org/10.1016/j.caeai.2021.100025>
- Zhang, W., & Tang, J. (2021). Teachers' TPACK Development: A Review of Literature. *Open Journal of Social Sciences*, 9(7), 367–380. <https://doi.org/10.4236/jss.2021.97027>

3 Digitale Unterrichtspraxis an kaufmännischen Schulen in der Corona-Pandemie - Eine Momentaufnahme

[Digital Teaching Practices in Commercial Schools in the Corona Pandemic - A Snapshot]

Study 1 was published in December 2023 in ‚Zeitschrift für Berufs- und Wirtschaftspädagogik‘. The publication is available at <https://doi.org/10.25162/zbw-2023-0021>. Below it is reproduced with permission from Franz Steiner Verlag.⁶

3.1 Abstract

Kurzfassung: Die Schulschließungen während der Corona-Pandemie führten zu großen Herausforderungen, insbesondere durch die überstürzte Umstellung auf Fernunterricht (Emergency-Remote-Teaching). Vor diesem Hintergrund beschreiben wir die digitale Unterrichtspraxis an kaufmännischen beruflichen Schulen während der Pandemie auf Basis der Angaben von 305 Lehrkräften. Häufig genutzte digitale Tools waren Lernplattformen, Videokonferenztools und Cloud-Lösungen. Auf Präsentationssoftware und Abstimmungstools wurde dagegen weniger häufig zurückgegriffen. Bei der Charakterisierung des Unterrichts auf Basis des SAMR-Modells (Puentedura, 2006) wurden die meisten digitalen Tools den Stufen Substitution und Augmentation zugeordnet, wohingegen Modification und Redefinition selten adressiert wurden. Lehrkräfte äußerten die Erwartung, dass digitale Tools den Präsenzunterricht auch zukünftig ergänzen.

Schlagworte: Digitalisierung, Digitales Lernen, Kaufmännische Schulen, Software, Lehrpersonen

Abstract: The school closures during the COVID-19 pandemic posed significant challenges, particularly due to the rushed transition to remote teaching (Emergency Remote Teaching). In this context, we describe the digital teaching practices at commercial vocational schools during the pandemic based on input from 305 teachers. Frequently used digital tools included learning platforms, video conferencing tools, and cloud solutions. However, presentation software and polling tools were used less frequently. When characterizing the teaching based on the SAMR

⁶ Note: The chapter headings, figures, and tables of the published version have been numbered according to the numbering of this thesis. Minor changes or deviations from the published paper publication are possible. Spelling and grammar have been adjusted to American English standards. Any misspellings and typographical errors detected in the course of a further thorough proofreading have been corrected.

model (Puentedura, 2006), most digital tools were classified under the Substitution and Augmentation levels, while Modification and Redefinition were seldom addressed. Teachers expressed the expectation that digital tools would continue to complement in-person instruction in the future.

Keywords: Digitization, Digital Learning, Vocational Schools, Software, Teachers

3.2 Emergency-Remote-Teaching in der Corona-Pandemie

Die Bedeutung der digital gestützten Unterrichtsgestaltung hatte in Zeiten der Corona-Pandemie enorm zugenommen. Innerhalb kürzester Zeit musste die bisherige Unterrichtspraxis an veränderte Lehr- und Lernbedingungen angepasst werden. Durch die Schulschließungen wurde der Fernunterricht – zumindest temporär – zur vorherrschenden Lehrform. Diese befristete Umstellung der Instruktionsform (i. S. einer „Lehrstrategie zur Steuerung des systematischen Lernens“, Weinert, 1999, S. 33) aufgrund einer temporären Krise stellte Lehrende und Lernende vor enorme Herausforderungen. Hodges et al. (2020) sprechen in diesem Zusammenhang von „Emergency-Remote-Teaching“. So erschwerten beispielsweise die unzureichende technische Infrastruktur sowie lediglich gering ausgeprägte digitale Kenntnisse die Situation (Huber et al., 2020). Auch fehlende didaktische Konzepte spielten hier eine Rolle (Sloane, 2020). Allerdings wird der Krise auch eine beschleunigende Wirkung hinsichtlich des digitalen Wandels von Lehr- und Lernprozessen zugeschrieben (Dilger, 2021), und es wird vermutet, dass die sich in der Pandemie stellenden Herausforderungen an Schulen bislang ungenutztes Potential offenlegen. Hindernisse und Hemmschwellen wurden abgebaut, und alles in allem kann eine gestiegene Bereitschaft zur digitalen Gestaltung von Lehr-Lernprozessen ausgemacht werden (Huber et al., 2020). Darüber hinaus kann Lernen mit digitalen Tools (einen didaktisch sinnvollen Einsatz vorausgesetzt) einen positiven Effekt auf die Lernergebnisse sowie die Einstellungen der Lernenden nach sich ziehen (für den mathematisch-naturwissenschaftlichen Unterricht siehe bspw. die Meta-Analyse von Hillmayr et al., 2020). Es werden positive Effekte berichtet, wobei die Effektstärken je nach Art des Tools variieren. Im Schnitt zeigt sich ein mittlerer Effekt ($g = .65$) auf die Leistung und große Moderationseffekte auf die Wirksamkeit durch Lehrkräfteschulungen ($g = .84$). Allerdings sind diesbezüglich auch Herausforderungen wie die lernförderliche Ausgestaltung und individuelle Passung zu nennen (Herzig, 2007).

Im Unterschied zu allgemeinbildenden Schulen (z. B. Wößmann et al., 2020, 2021) lassen sich für die Berufsbildung nur wenig Erkenntnisse zum digitalen Unterricht während der Coronapandemie ausmachen (Seifried, 2021). Über die konkrete Ausgestaltung der (Präsenz-) Lehre,

von (Teil-) Schulschließungen bis hin zu Wechselunterricht ist wenig bekannt, da die Einschränkungen von Präsenzunterricht den jeweiligen regionalen Anpassungen unterlagen. Vor diesem Hintergrund ist es von Interesse, wie sich die Nutzung digitaler Technologien in der kaufmännischen Unterrichtspraxis während der Schulschließungen gestaltete und welche der in dieser Phase erarbeiteten Errungenschaften dauerhaft bestehen werden. Zur Charakterisierung des Unterrichts greifen wir auf das in der Bildungspraxis geläufige SAMR-Modell (Puentedura, 2006, 2014) zurück, welches den Einsatz digitaler Tools auf vier Stufen verortet (Substitution, Augmentation, Modification, Redefinition). Im vorliegenden Beitrag werden in der Tradition der deskriptiven Unterrichtsforschung (siehe u.a. Seifried, 2008) drei Fragestellungen bearbeitet:

- (1) Wie gestaltete sich die digitale Unterrichtspraxis an kaufmännischen Schulen (Einsatz von digitalen Tools) während der Corona-Pandemie?
- (2) Wie lässt sich die digitale Unterrichtspraxis an kaufmännischen Schulen charakterisieren (unter Rückgriff auf das SAMR-Modell)?
- (3) Über welche Erwartungen bezüglich der Unterrichtspraxis nach der Corona-Pandemie berichten Lehrkräfte?

Der vorliegende Beitrag ist wie folgt gegliedert: In Abschnitt 3 werden Modelle und Ansätze zum unterrichtlichen Einsatz von digitalen Tools skizziert. Hierbei gehen wir insbesondere auf das SAMR-Modell ein. Anschließend werden das methodische Vorgehen dargestellt (Abschnitt 4) sowie die Ergebnisse einer Befragung von Lehrkräften an kaufmännischen Schulen in Baden-Württemberg zum Einsatz digitaler Tools berichtet (Abschnitt 5). Abschließend diskutieren wir die Ergebnisse und geben einen Ausblick auf zukünftige Entwicklungen.

3.3 Digitale Tools im Unterricht

Digitaler Unterricht (auch Distance Education; Distance Learning; Remote Education) wird definiert als „planned learning that normally occurs in a different place from teaching and as a result requires special techniques of course design, special instructional techniques, special methods of communication by electronic and other technology, as well as special organizational and administrative arrangements“ (Moore & Kearsley, 1996, S. 2). Distance Learning ist dabei nicht deckungsgleich zu E-Learning (Wheeler, 2012): Clark und Mayer (Clark & Mayer, 2016, S. 30) beispielsweise fassen *E-Learning* als „instruction delivered on a digital device that is intended to support learning“. E-Learning (auch *technology-enhanced-learning*, *digital learning*) meint dabei den Einsatz technologischer Werkzeuge (digitale Medien und Tools) zur Unterstützung des Lernens (Wheeler, 2012). Diese breite Definition umfasst asynchrone und synchrone Lehr-Lernsettings ebenso wie verschiedene Distributionsmittel (Laptop, Tablet, etc.).

Der Einsatz digitaler Tools (worunter in diesem Beitrag neben den Tools auch digitale Medien wie E-Books oder Erklärvideos gezählt werden) ist kein Selbstzweck, sondern muss einen lernförderlichen Mehrwert bieten (Scheiter, 2021). Empirische Befunde zu möglichen Effekten liegen seit geraumer Zeit vor und werden unter den Schlagworten *technology-enhanced learning* bzw. *technology-enhanced teaching* diskutiert. Dabei zeichnet sich die ältere Forschungsperspektive (Lernen mit digitalen Medien) durch einen Fokus auf Lernprozesse und -produkte in Abhängigkeit von medialen und individuellen Merkmalen aus. Demgegenüber richtet die jüngere Forschungsperspektive (Lehren mit digitalen Medien) den Blick auf die Gestaltung des Unterrichts und die hierfür notwendigen Kompetenzen des Lehrpersonals (Scheiter, 2021). In jüngerer Zeit vorgelegte Metaanalysen verweisen auf die Möglichkeit des lernförderlichen Einsatzes digitaler Tools (für den Elementarbereich siehe Chauhan, 2017; für die mathematisch-naturwissenschaftliche Domäne siehe Hillmayr et al., 2020). Viele Befunde beziehen sich jedoch auf den Einsatz von Technologien im Präsenzunterricht und lassen keine Rückschlüsse auf die Effekte von Fernunterricht zu. Studien zu den Effekten von Fernunterricht verweisen vor allem auf die Notwendigkeit der Kombination von Fern- und Präsenzunterricht (Bernard et al., 2014; Means et al., 2010). Aus einer schulartübergreifenden Befragung von Schülerinnen und Schülern in Baden-Württemberg zur ersten Phase des pandemiebedingten Fernunterrichts (Wacker et al., 2020) geht zudem hervor, dass die Lernenden – neben einigen Problembereichen wie die fehlende Kommunikation/Rückmeldung/Unterstützung oder Probleme mit dem selbstständigen Arbeiten – auch Vorteile (z. B. Flexibilisierung/Individualisierung) sehen (einen Überblick über den Forschungsstand bieten Helm et al., 2021). Entsprechend herausfordernd gestaltet sich der lernwirksame Einsatz digitaler Medien.

Unstrittig ist, dass sich mit dem Einsatz digitaler Medien vielfältige Optionen für Lehrende und Lernende eröffnen (Howe & Knutzen, 2013). Digitale Tools ermöglichen nicht nur den Austausch von Informationen, sondern weiterführend auch Kommunikation und Kollaboration sowie die Veranschaulichung und Simulation von Abläufen. Zudem werden seit geraumer Zeit auch berufs- bzw. domänenspezifische Tools implementiert (bspw. Enterprise-Resource-Planning Systeme, siehe dazu Spener & Schumann, 2021). Darüber hinaus können digitale Medien für Assessments genutzt werden oder die Lernenden bei der Reflexion ihrer Kompetenzen unterstützen (für eine praxisorientierte Klassifizierung von digitalen Tools siehe auch Busch, 2020 sowie Howe & Knutzen, 2013). Der lernwirksame Einsatz digitaler Instrumente setzt schließlich auch entsprechende Kompetenzen der Lehrpersonen voraus (z.B. Seufert et al., 2018). Mishra und Koehler (2006) verweisen diesbezüglich im Rahmen des Technological Pedagogical Content Knowledge Frameworks (TPACK) auf das komplexe Zusammenspiel der drei Wissensbereiche Content (CK), Pedagogy (PK) und Technology (TK). Darüber hinaus ermöglichen

Rahmenwerke wie DigCompEdu (Redecker, 2017, Europäische Union) eine Verortung beruflicher, pädagogischer und didaktischer Kompetenzen der Lehrenden sowie Kompetenzen der Lernenden.

Zur Klassifizierung der verschiedenen Einsatzmöglichkeiten von digitalen Tools beim Lehren und Lernen werden häufig Modelle des technology-enhanced teachings wie SAMR (Substitution, Augmentation, Modification, Redefinition; Puentedura, 2006, 2014) oder RAT (Replacement, Amplification and Transformation; Hughes et al., 2006) herangezogen. Diese Modelle postulieren eine Unterrichtsentwicklung über verschiedene Stufen (von einem Ersatz herkömmlicher Praktiken bis hin zu einer fundamentalen Neugestaltung des Unterrichts Scheiter, 2021). Dabei hat insbesondere das SAMR-Modell Bekanntheit erlangt. Der Einsatz digitaler Medien wird hier auf zwei Ebenen (Enhancement und Transformation) bzw. vier Stufen verortet (siehe Figure 3-1), wobei Stufe 1 und 2 der Ebene des Enhancements und Stufe 3 und 4 der Ebene der Transformation zugerechnet werden. Auf der Stufe der Substitution werden analoge Medien durch digitale Tools ersetzt, es findet jedoch keine funktionale Veränderung statt. Dies ist beispielsweise der Fall, wenn statt auf ausgedruckte Lernmaterialien auf eine digitale Version zurückgegriffen wird. Die funktionale Erweiterung durch den Einsatz digitaler Tools wird als Augmentation bezeichnet. Exemplarisch kann hier die Verwendung von Rechtschreibprüfungen in Textverarbeitungsprogrammen angeführt werden. Auf der Modifikations-Stufe wird durch die digitale Unterstützung eine Neugestaltung von Aufgaben möglich. Dies umfasst beispielsweise das gemeinsame Arbeiten anhand kollaborativer Tools. Neuartige Aufgabenstellungen oder Unterrichtsformate, die ohne den Einsatz neuer Technologien nicht umsetzbar wären, werden schließlich als Redefinition bezeichnet. Auf dieser Stufe wäre zum Beispiel die Nutzung von komplexen Lehr-Lern-Arrangements wie Simulationen zu verorten.

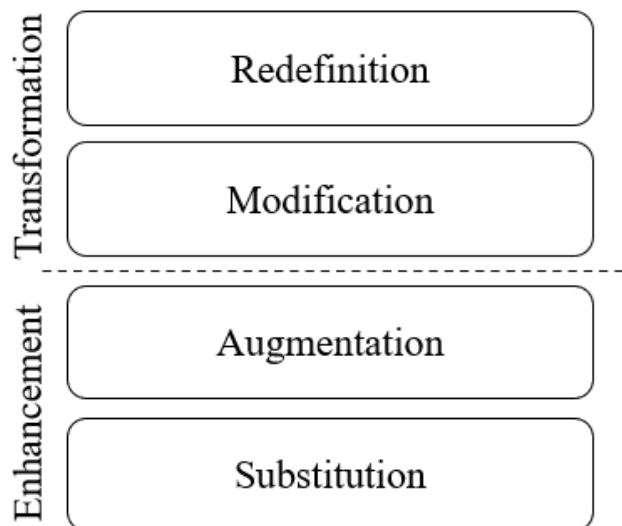


Figure 3-1: Stufen des SAMR-Modells (adaptiert von Puentedura, 2006).

Das SAMR-Modell gilt als plausibel und gut nachvollziehbar, wird aber auch kritisch betrachtet. Hamilton et al. (2016) kritisieren zum einen den fehlenden Kontextbezug des Modells, da zentrale Einflussvariablen wie technologische Infrastruktur, Ressourcen und individuelle Bedürfnisse von Lernenden sowie die Kompetenzen der Lehrkräfte unberücksichtigt bleiben. Durch die Vernachlässigung des spezifischen Unterrichtskontexts werde das SAMR-Modell der Komplexität fachbezogener Lehr-Lernsituationen nicht gerecht. Zudem wird die rigide Struktur des Modells bemängelt. Die Stufen des Modells bilden einen pragmatischen Determinismus ab. Die hierarchische Struktur der verschiedenen Stufen impliziere somit eine Wertigkeit des Einsatzes von Technologien, welche die pädagogische Komponente in den Hintergrund rücken lässt. Schließlich wird die Fokussierung auf digitale Tools und Medien kritisiert. Besser wäre es, auf die hierdurch angeregten Prozesse abzustellen. Das SAMR-Modell ziele somit lediglich auf die Veränderung von Unterrichtsaktivitäten ab und nicht auf eine Steigerung der Qualität von Lernprozessen.

Ungeachtet dieser berechtigten Kritikpunkte wird das SAMR-Modell immer wieder zur Beschreibung der digitalen Unterrichtspraxis herangezogen. In einer Studie zur Klassifizierung des Biologieunterrichts (Kramer et al., 2019) wurden etwa 70% der eingesetzten digitalen Medien der Ebene der Substitution und 10% der Ebene Augmentation zugeordnet. Modification und Redefinition war nicht beobachtbar, und für rund 20% des Medieneinsatzes war keine Einordnung in SAMR möglich. Ähnliche Befunde zeigen sich in einer Studie, die den Einsatz der Lernplattform Moodle in Österreichs Schulen während des Distance Learnings analysiert (Sch-

renk, 2021). Obwohl die verfügbaren Funktionen und Tools der Plattform die Möglichkeit eröffnen, alle vier Stufen des SAMR-Modells zu adressieren, zeigt sich, dass die Plattform hauptsächlich für den Austausch von Dateien und Aufgaben genutzt wurde. Es wird deutlich, dass die Bereitstellung bzw. Verfügbarkeit von digitalen Möglichkeiten eine notwendige, aber keine hinreichende Bedingung für einen lernförderlichen Einsatz darstellt: „Lehrer:innen, die erstmalig mit digitalen Technologien arbeiten, tendieren dazu, herkömmliche Lehrmuster und -materialien 1:1 aufs Digitale zu übertragen“ (Schrenk, 2021, S. 59). Aus didaktischer Perspektive ist es vielmehr notwendig, die aktive Auseinandersetzung der Lernenden mit den Inhalten zu fördern oder elaboriertes Feedback anzubieten (Schmid et al., 2009; Tamim et al., 2011; Van der Kleij et al., 2015). Schließlich sind auch rechtliche Aspekte (u. a. Datenschutz) zu klären. Es ist daher unmittelbar nachvollziehbar, dass seitens der Lehrkräfte Vorgaben bezüglich des Einsatzes digitaler Tools gewünscht werden (Ifenthaler & Delcker, 2021).

3.4 Methodisches Vorgehen

Datenerhebung und Stichprobe

Um mehr über die Unterrichtsgestaltung an kaufmännischen Schulen während der Corona-Pandemie zu erfahren, wurde eine Onlinebefragung von Lehrkräften durchgeführt. Diese fand Anfang April 2021 über einen Zeitraum von zwei Wochen und somit zu Zeiten erneuter (Teil-)Schulschließungen bzw. Wechselunterrichts während der dritten Welle der Corona-Pandemie (siehe Bujard et al., 2021, S. 10) statt.

Im Vorfeld der Datenerhebung wurden Schulleiter:innen sämtlicher kaufmännischer Schulen in Baden-Württemberg kontaktiert und gebeten, die Informationen und Zugangsdaten zum Fragebogen an das Kollegium weiterzuleiten. Die Teilnahme war freiwillig und wurde nicht vergütet. Eine Rücklaufquote lässt sich aufgrund des gewählten Distributionsweges nicht ermitteln. Insgesamt haben 387 Lehrkräfte an der Umfrage teilgenommen. In die Auswertung gingen ausschließlich vollständig ausgefüllte Fragebögen ein ($n = 305$; 79% vollständig ausgefüllte Fragebögen).

Die 305 antworteten Lehrkräfte (68% weiblich, 30% männlich, 2% keine Angabe) waren zum Zeitpunkt der Erhebung durchschnittlich 45 Jahre alt ($M = 44.7$, $SD = 11.0$). Mehr als die Hälfte (55%) der Teilnehmenden hat eine dem Studium vorangegangene kaufmännische Berufsausbildung abgeschlossen. Im Schnitt unterrichten die Lehrkräfte seit über 15 Jahren ($M = 15.3$,

$SD = 10.05$, $Max = 42$ Dienstjahre) und sind überwiegend mit vollem Deputat beschäftigt (Vollzeit 59%, Teilzeit 37%, Referendariat 4%).⁷ Die meisten Lehrkräfte sind als Fachlehrkraft tätig (Fachlehrkraft 69%, Fachbereichsleitung 8%, Schulleitung 6%, IT/Support-Lehrkraft 3%, Sonstiges 7%, keine Angabe 7%) und unterrichten überwiegend an einem beruflichen Gymnasium ($n = 189$) und an der Berufsschule ($n = 159$). Ein kleinerer Teil der Lehrkräfte unterrichtet überwiegend an einem Berufskolleg ($n = 89$) oder in der Berufsfachschule ($n = 74$). Nur in Ausnahmefällen sind die Befragten überwiegend an einer Fachschule ($n = 6$) oder Berufsoberschule ($n = 3$) tätig. Die Lehrenden unterrichten überwiegend wirtschaftswissenschaftliche Inhalte (BWL; $n = 179$; VWL; $n = 101$). Darüber hinaus werden Englisch ($n = 50$), Deutsch ($n = 48$) und Mathematik ($n = 39$) sowie Geschichte bzw. Gemeinschaftskunde ($n = 44/39$) genannt.

Befragungsinstrument

Der eingesetzte Fragebogen diente dem Ziel der Beschreibung der digitalen Unterrichtspraktiken und beinhaltet Items zu unterschiedlichen Bereichen: technische Ausstattung und Unterstützung, Anteil des asynchronen Unterrichts, Nutzung digitaler Tools, Digitalisierungsgrade nach SAMR, Erwartungen an die zukünftige Unterrichtspraxis, Lehr-Lern-Überzeugungen sowie Selbstbewusstsein im Umgang mit Computern. Falls möglich, wurde auf bewährte Instrumente zurückgegriffen und diese entweder direkt übernommen oder für die Zielgruppe adaptiert. Einige der Items stellen Eigenentwicklungen dar.

Technische Ausstattung/Unterstützung und A-Synchronität des Unterrichts

Es wurden drei Items in Bezug auf die technische Ausstattung/Unterstützung (Beispiel-Item: „An meiner Schule werden technische Fertigkeiten zum Umgang und zur Nutzung digitaler Medien vermittelt.“) eingesetzt (eigene Items, siehe dazu auch die Analyse von Kerres & Kalz, 2003). Die Erfassung erfolgte über eine fünfstufige Likert-Skala von 1 = trifft überhaupt nicht zu bis 5 = trifft voll und ganz zu. Die Reliabilität der Items zur technischen Ausstattung ist als akzeptabel zu bewerten ($\alpha = .74$). Zusätzlich wurde die (A-)Synchronität der Unterrichtspraxis erfragt („Wie hoch schätzen Sie den prozentualen Anteil der asynchronen Unterrichtspraxis?“, Skala von 0 bis 100).

⁷ Der Anteil an Lehrerinnen mit verringertem Stundendeputat ist im Vergleich zur Gruppe der männlichen Lehrer deutlich höher (49% vs. 13% Teilzeit).

Verwendung digitaler Tools

Die Verwendung von digitalen Tools (Auswahl der Tools in Anlehnung an Busch, 2020 sowie Howe & Knutzen, 2013; eigene Items, siehe dazu auch die Angaben zur Nutzung digitaler Medien von Drossel et al., 2019) wurde durch zehn Items erfasst (z. B.: „Wie häufig verwenden Sie Lernplattformen bzw. Kursmanagementsysteme zur Verwaltung von Klassen/Organisation von Unterricht?“). Zur Messung der Nutzungsfrequenz wurde eine sechsstufige Skala von 1 = nie bis 6 = mehrmals täglich herangezogen. Die Reliabilität für die Items zur Verwendung von Tools ist als gut zu kennzeichnen ($\alpha = .80$).

Stufen der Digitalisierung des Unterrichts

Es wurden dreizehn Items zur Beschreibung der Digitalisierung des Unterrichts entwickelt (in Anlehnung an: Drugova et al., 2021; Lubega et al., 2014; Puentedura, 2006; Wahyuni et al., 2020). Dem SAMR-Modell folgend wurden zwei Ebenen unterschieden: (1) Die Ebene der Verbesserung (Enhancement: sechs Items; Aggregation der Stufen Substitution und Augmentation) und (2) die Ebene der Umgestaltung (Transformation: sieben Items; Aggregation der Stufen Modification und Redefinition). Ein Beispiel aus dem Bereich der Verbesserung des Unterrichts durch den Einsatz digitaler Tools lautet: „Ich weise meine Schülerinnen und Schüler auf die Rechtschreibkorrektur in Office-Anwendungen (z.B. MS Word, OpenOffice) hin“. Als Beispiel für die Umgestaltung des Unterrichts durch den Einsatz digitaler Tools ist zu nennen: „Ich ermutige meine Schülerinnen und Schüler, auf Open Education Resources (z.B. Lernvideos, MOOCs) eigenständig zuzugreifen“. Die Erfassung erfolgte über eine vierstufige Likert-Skala von 1 = stimme überhaupt nicht zu bis 4 = stimme voll und ganz zu. Zur Sicherung der Konsistenz wurde das Instrument von ursprünglichen sechzehn auf dreizehn Items reduziert, wodurch sich die Reliabilitäten verbesserten. Vereinzelt wurden Items aufgrund unzureichender Trennschärfe (Itemtrennschärfe $< .3$) aus der weiteren Analyse entfernt. Die Reliabilitäten der Items sind letztlich als (noch) akzeptabel zu bewerten (Gesamtskala SAMR: $\alpha = .79$; Subskala Enhancement: $\alpha = .71$, Subskala Transformation: $\alpha = .68$, siehe hierzu die nachstehenden Ausführungen).

Faktoranalytische Betrachtungen zur Prüfung der Güte des Instruments verweisen jedoch auf Probleme, welche letztlich nicht zufriedenstellend aufgelöst werden konnten. Folgt man dem SAMR-Modell, sollten eigentlich vier Stufen zu unterscheiden sein. Eine konfirmatorische Analyse (ML-Schätzung) zur Prüfung der Vier-Faktoren-Struktur führt mit Blick auf die üblichen Grenzwerte (siehe Schermelleh-Engel et al., 2003) zu unbefriedigenden Ergebnissen (RMSEA $.092 > .08$; CFI $0.815 < .90$). Eine ergänzend durchgeführte explorative Faktorenanalyse verweist auf eine Zwei-Faktorenlösung, nämlich die Unterscheidung von Enhancement

und Transformation, die sich aus modelltheoretischer Perspektive auch begründen lässt. Vor diesem Hintergrund beschränken wir uns vornehmlich auf eine Analyse der Unterrichtspraxis auf Item-Ebene und eine deskriptive Darstellung der Ergebnisse. Angesichts der skizzierten Problemlage wurde dann auch auf tiefergehende statistische Auswertungen verzichtet.

Erwartungen zur zukünftigen digitalen Gestaltung des Unterrichts

Mittels eines Freitext-Items wurde nach den Erwartungen bezüglich des zukünftigen digitalen Unterrichts gefragt: „Was bleibt von der Digitalisierung des Unterrichts, auch nach der Coronakrise, längerfristig bestehen? Was denken Sie?“

Auswertung

In die Datenauswertung gingen ausschließlich vollständig ausgefüllte Fragebögen ein. Die quantitative Datenauswertung erfolgte mit R. Die Auswertung der Textantworten der Lehrkräfte erfolgte anhand einer qualitativen Inhaltsanalyse mittels MAXQDA. Hierzu wurde auf einen induktiv entwickelten Kodierleitfaden zurückgegriffen, welcher die Kategorien „Unterrichtsgestaltung“, „Rahmenbedingungen“, „Rückkehr zur prepandemischen Unterrichtspraxis“ sowie „Sonstiges“ umfasst. Innerhalb der beiden erstgenannten Kategorien wurden weitere Subcodes zur Differenzierung der Antworten angelegt. Um die Zuverlässigkeit des Kodiersystems zu prüfen, wurde das Datenmaterial von zwei Personen kodiert und anschließend die Inter-coder-Reliabilität bestimmt. Der ermittelte Kappa-Wert von 0.81 verweist auf eine zufriedenstellende Übereinstimmung. Insgesamt ergaben sich 402 Kodierungen. Tabelle 3-1 zeigt exemplarisch die Subcodes für die Kategorie „Unterrichtsgestaltung“:

Table 3-1: Ausschnitt aus dem Kodierleitfaden zur Kategorie Unterrichtsgestaltung.

Subcodes	Kodierregel	Ankerbeispiel
1.1 Einsatz digitaler Tools	Wird kodiert, wenn der Einsatz digitaler Tools in der Unterrichtspraxis thematisiert wird (auch wenn lediglich die Nennung eines Toolnamens oder Einsatzzwecks erfolgt).	z. B. Aussagen bezüglich Lernplattformen, Kursmanagementsystemen, Lernvideos, Quiz; „Bereitstellung von Materialien auf Moodle“; „methodische (digitale) Alternativen im Präsenzunterricht“
1.2 Einsatz digitaler Endgeräte	Wird kodiert, wenn der Einsatz von digitalen Endgeräten in der Unterrichtspraxis thematisiert wird.	z. B. „Recherche mit dem Handy bzw. Endgerät durch die Schüler“; „Tabletnutzung“; „Hausaufgaben PC-gestützt“
1.3 Lernprozesse	Wird kodiert, wenn (veränderte) Lernprozesse oder Arbeitsweisen der Lernenden thematisiert werden.	z. B. „Eigenständigkeit der Schüler“; „selbstständige und kollaborative Arbeitsweisen“; „mehr asynchrone Lernmöglichkeiten“
1.4 Assessment	Wird kodiert, wenn Formen der Leistungsüberprüfung thematisiert werden.	z. B. „Alternative Leistungsfeststellungen (z. B. Podcast erstellen)“
1.5 Sonstiges	Wird kodiert, wenn eine sonstige Aussage zur Unterrichtsgestaltung vorliegt, die sich nicht in die vorherigen Kategorien einordnen lässt.	z. B. „Unterrichtsvorbereitung über mehrere Bildschirme“; „der Unterricht an der Tafel geht in Zukunft gegen Null“

3.5 Empirische Befunde

Nutzung digitaler Tools in der Unterrichtspraxis

Zunächst lässt sich festhalten, dass der digitale Unterricht nur zu geringen Teilen asynchron stattfand ($M = 33\%$ der Unterrichtszeit, $SD = 24.1\%$) und größtenteils im synchronen Format durchgeführt wurde. Mit Blick auf die technische Ausstattung und organisationale Unterstützung gaben 54% der Lehrkräfte an, dass ein leistungsfähiger Internetanschluss an der Schule zur Verfügung steht. Knapp ein Viertel (23%) der Lehrkräfte erklärte hingegen, dass kein ausreichender Internetanschluss vorhanden sei (weitere 23% der Lehrkräfte stimmen nur teilweise zu). Gut die Hälfte (53%) der Lehrkräfte bestätigte, dass an der Schule technische Fertigkeiten zum Umgang und zur Nutzung neuer digitaler Medien vermittelt werden (16% stimmen nicht zu; 30% nur teilweise). Die Lehrkräfte gaben auch mehrheitlich an, dass sie von ihrer Schule bei der kompetenten und reflektierten Nutzung neuer Medien unterstützt werden (41% der Lehrkräfte stimmen hier zu, 26% stimmen nicht zu; 31% nur teilweise zu). Ein konträres Bild zeichnet sich bei der Frage nach dem Einsatz moderner Technologien ab (beispielsweise Produktionsanlagen (4.0), Business Intelligence Tools, Virtual Reality, etc.). Lediglich 15% der Befragten stimmten hier zu (69% trifft eher nicht zu; 16% trifft teilweise zu).

Bezüglich der Häufigkeit der Nutzung der Tools lassen sich drei Abstufungen ausmachen: (1) Häufig verwendete Tools, die überwiegend mehrmals (d.h. mehr als einmal) die Woche oder

häufiger eingesetzt wurden, (2) gelegentlich verwendete Tools, die von einer Hälfte der Lehrkräfte etwa einmal die Woche oder seltener und von der anderen Hälfte mehrmals die Woche oder häufiger eingesetzt wurden, sowie (3) selten verwendete Tools, die überwiegend etwa einmal die Woche oder weniger eingesetzt wurden (siehe Figure 3-2). Zu den häufig verwendeten Tools gehören Plattformen für Videokonferenzen, Lernplattformen bzw. Kursmanagementsysteme (Learning Management Systeme) zur Klassenverwaltung und Organisation des Unterrichts sowie Clouds zum Teilen von Dateien. Videokonferenzen wurden von 97% der Lehrkräfte häufiger eingesetzt (77% mehrmals täglich). Lernplattformen wurden von 95% (79% täglich) und Cloud-Speicher von 86% häufiger genutzt. Unter den gelegentlich verwendeten Werkzeugen finden sich Tools zum Abspielen von Videos (online und offline) und Präsentationsprogramme. Videos wurden dabei von 58% der Lehrpersonen gelegentlich (34% mehrmals die Woche) und von 42% eher selten (also einmal die Woche bis nie) im digitalen Unterricht eingesetzt. Präsentationsprogramme wurden von 57% der Lehrpersonen gelegentlich (davon 28% mehrmals täglich) und von 43% der Lehrkräfte seltener eingesetzt (24% weniger als einmal die Woche). Noch seltener eingesetzt wurden Tools zur gemeinsamen schriftlichen Diskussion, zur Erstellung und Durchführung von Umfragen und Abstimmungen, zur gemeinsamen Bearbeitung von Dateien, zur Erstellung und Durchführung von Quizen und Tests, und Tools zum Sammeln und Strukturieren von Ideen. Eine gemeinsame schriftliche Diskussion über digitale Tools fand bei 65% der befragten Lehrkräfte überwiegend eher unregelmäßig statt, d.h. etwa einmal die Woche oder seltener (davon 33% nie). Abstimmungen und Umfragen wurden von 72% der Lehrkräfte ebenfalls selten eingesetzt (34% weniger als einmal die Woche). Von den Befragten im digitalen Unterricht eher selten genutzt wurde das gemeinsame Bearbeiten von Dateien (74%, davon 42% nie), die Durchführung von Tests und Quizen (80%, davon 34% nie) und das Sammeln und Strukturieren von Ideen (89%, davon 45% nie).

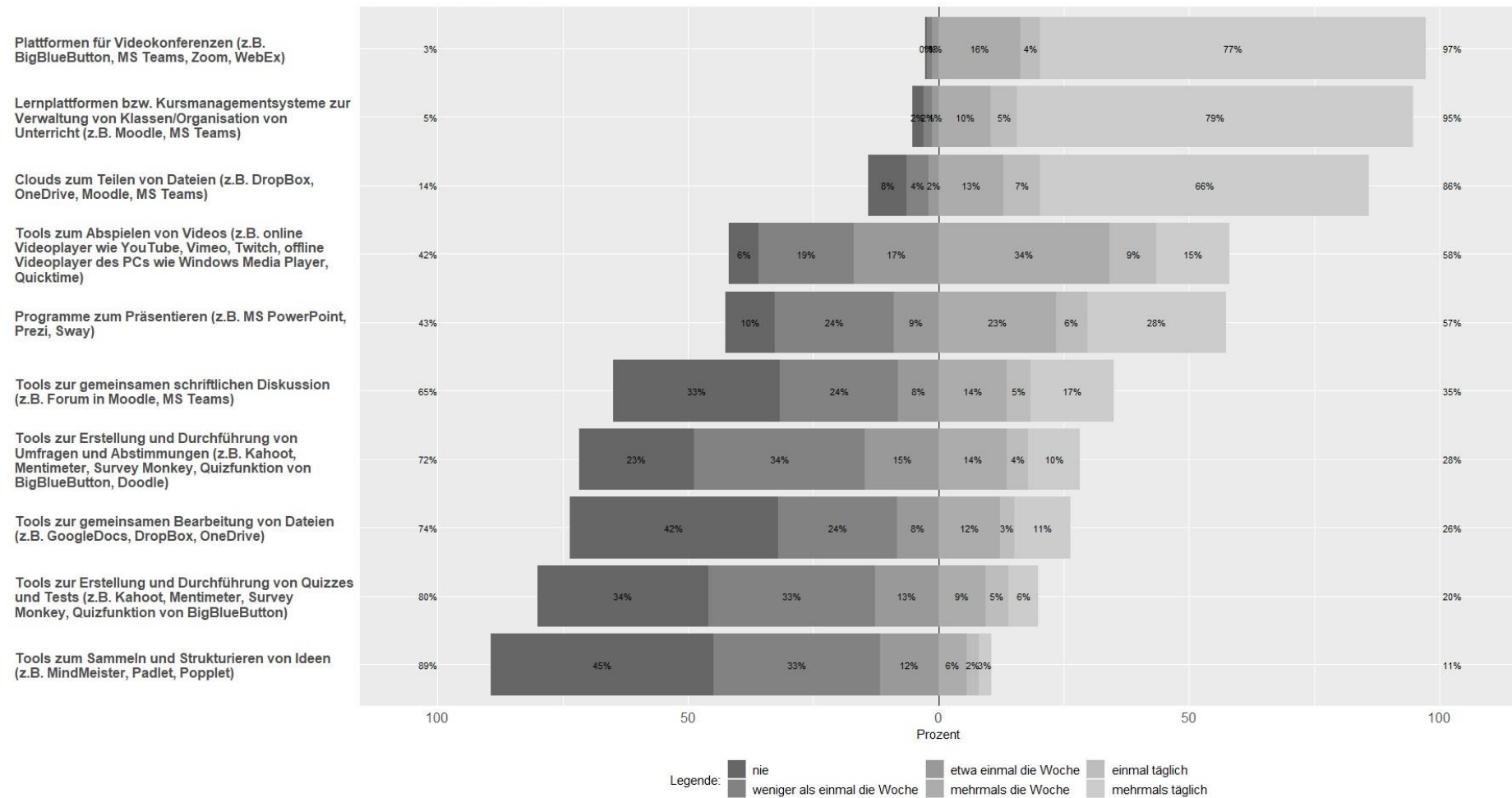


Figure 3-2: Nutzung von digitalen Tools (eigene Darstellung).

Charakterisierung der Unterrichtspraxis unter Rückgriff auf das SAMR-Modell

Zur Charakterisierung der digitalen Unterrichtspraxis kann – wie bereits ausgeführt – in Anlehnung an das SAMR-Modell zwischen den Ebenen „Verbesserung“ (Enhancement) und „Umgestaltung“ (Transformation) unterschieden werden. Es zeigt sich, dass Lehrkräfte vermehrt digitale Tools einsetzen, die zur Stufe der Verbesserung der Unterrichtspraxis zählen (siehe Figure 3-3). Die mittlere Zustimmung beträgt $M = 2.49$ ($SD = .63$). Die Darbietung von Unterrichtsstoff über Präsentationstools (64% Zustimmung), die Anregung, bei Verständnisproblemen auf Online-Nachschlagewerke (63% Zustimmung) zurückzugreifen, sowie die Lernenden aktiv Informationen über Suchmaschinen (58% Zustimmung) ermitteln zu lassen, sind häufig genutzte Lehrstrategien. Eine Mehrheit lehnt die Verwendung von Online-Ressourcen als Ersatz für gedruckte Schulbücher ab. Ebenso stehen Lehrkräfte der gemeinsamen Bearbeitung von Aufgaben via Onlinedienste eher ablehnend gegenüber (jeweils über 70% Ablehnung).

Digitale Tools, die zu einer Neugestaltung der Unterrichtspraxis führen könnten, werden dagegen seltener eingesetzt (siehe Figure 3-4). Hier beträgt die mittlere Zustimmung 1.98 ($SD = .57$). Lehrkräfte stimmen zwar überwiegend zu, die Lernenden zur Nutzung kollaborativer Tools für Diskussionen und selbstständige Organisation von Gruppenarbeiten zu ermutigen (58% Zustimmung). Ein tendenziell eher ablehnendes Bild (Lehrkräfte stimmen überwiegend eher nicht zu) zeigt sich aber bezüglich der Ermutigung der Lernenden, selbstständig auf Open Educational Resources (OER) zuzugreifen (55% Ablehnung), sowie bei der Nutzung von Online Assessment Tools zur Evaluation des Lernstandes (60% Ablehnung). Über den Einsatz komplexer Lehr-Lern-Arrangements wie Büro-Simulationen (77% Ablehnung) oder die Nutzung von Learning Analytics (90% Ablehnung) zum Monitoring von Lernfortschritten wird nur selten berichtet. Ähnlich geringe Zustimmung findet auch die Aussage, dass Lernende eigene OER erstellen (81% Ablehnung) oder mit Serious Games oder intelligenten Tutorensystemen (94% Ablehnung) in Berührung kommen.

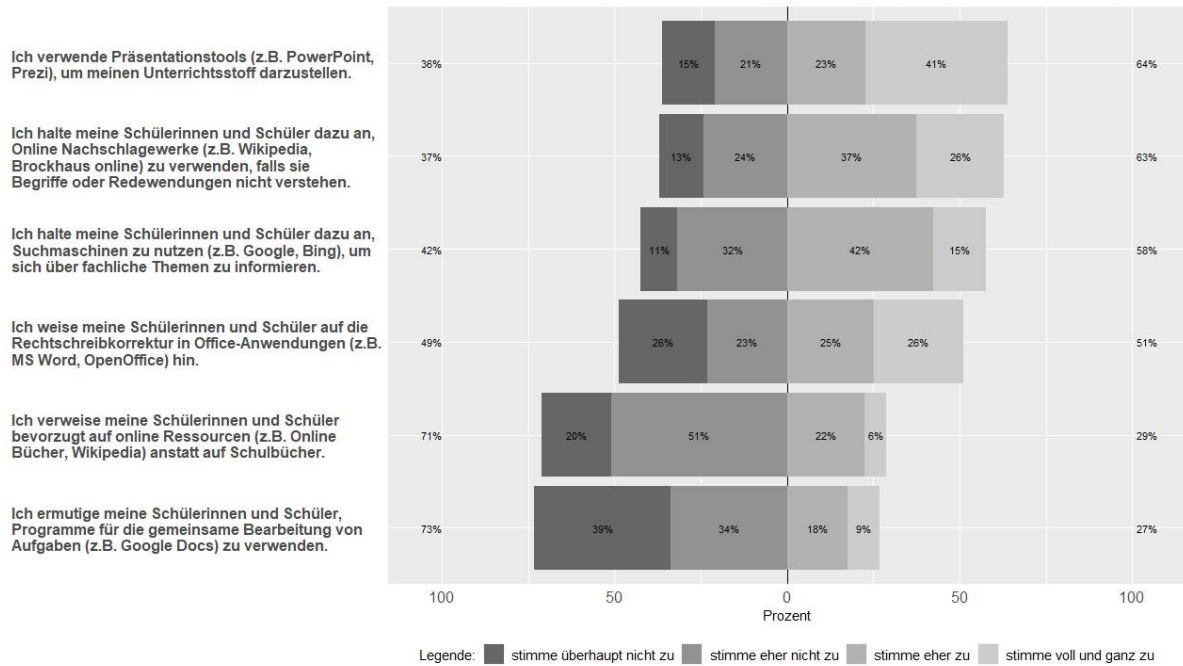


Figure 3-3: Zustimmung bzw. Ablehnung zu digitalen Tools der Enhancement Ebene (eigene Darstellung).

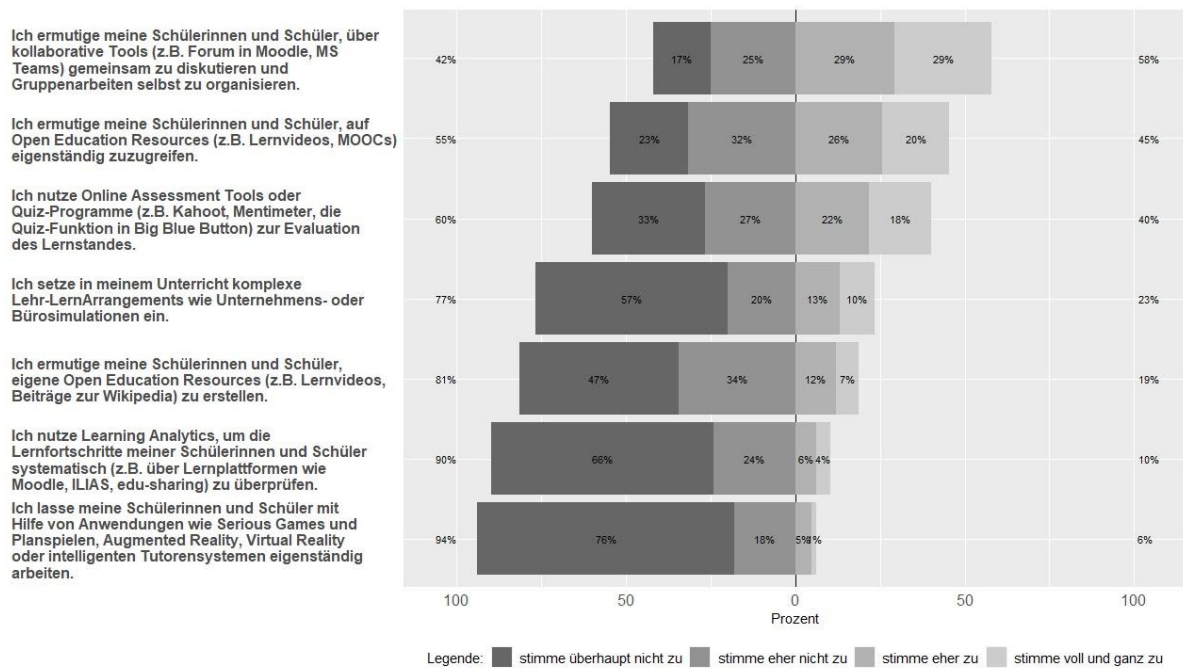


Figure 3-4: Zustimmung bzw. Ablehnung zu digitalen Tools der Transformation Ebene (eigene Darstellung).

Zusammenfassend lässt sich festhalten, dass auf der Ebene der Verbesserung (Enhancement) die Zustimmung stärker ausfällt als auf der Ebene der Neugestaltung des Unterrichts. Der Mittelwert für Enhancement ($M = 2.49, SD = .63$) unterscheidet sich signifikant ($t(304) = 15.155, p < .001$) vom Wert für Transformation ($M = 1.98, SD = .57$). Nach Cohen (1988) liegt mit $d = .87$ ein großer Effekt vor.

Erwartungen an die postpandemische Unterrichtspraxis

Bezüglich der erwarteten langfristigen Auswirkungen der Digitalisierung des Unterrichts auf die Unterrichtspraxis nach Corona lässt sich Folgendes festhalten: Die meisten Aussagen der befragten Lehrkräfte entfallen auf die *Unterrichtsgestaltung*, wobei der *Einsatz digitaler Tools* am häufigsten thematisiert wird (für eine vollständige Übersicht der angelegten Codes und deren Häufigkeit siehe Tabelle 3-2). Demnach geht ein Großteil der befragten Lehrkräfte davon aus, dass digitale Tools auch zukünftig eine Rolle spielen werden. Dies betrifft v. a. die Bereitstellung von Unterrichtsmaterialien und die Organisation von Lernunterlagen über Lernplattformen. Ferner werden die Möglichkeiten der Ergänzung des Präsenzunterrichts mit „methodischen (digitalen) Alternativen“ (Person 53, nachfolgend P) und beispielsweise der Einsatz von Lernvideos, Quiz oder Online-Nachschlagewerken genannt. Daneben wird auch von einem *Einsatz digitaler Endgeräte* im Präsenzunterricht ausgegangen. Insbesondere die Verwendung von Tablets für Recherchezwecke oder für ein digitales Dokumentenmanagement werden hier hervorgehoben. Hinsichtlich der Erwartungen an die Gestaltung der *Lernprozesse* dominiert die Auffassung, dass selbstreguliertes Lernen auch zukünftig von Bedeutung sein wird und Lernende sich vermehrt eigenständig Themen erarbeiten müssen.

Table 3-2: Erwartete langfristige Auswirkungen der pandemiebedingten Digitalisierung des Unterrichts auf die postpandemische Unterrichtspraxis (Anzahl der jeweiligen (Sub-)Codes).

(Sub-)Codes	#
Unterrichtsgestaltung	251
Einsatz digitaler Tools	174
Einsatz digitaler Endgeräte	34
Lernprozesse	15
Assessment	2
Sonstige Aussagen zur Unterrichtsgestaltung	26
Rahmenbedingungen	102
Kommunikation	64
Entwicklung digitaler Kompetenzen	15
Technische Ausstattung	13
Erwartungen an Lehrkräfte	7
Sonstige Rahmenbedingungen	3
Rückkehr zu prepandemischer Unterrichtspraxis	22
Sonstiges	27

Darüber hinaus werden bezüglich der *Rahmenbedingungen* für Schule und Unterricht verschiedene Auswirkungen thematisiert. Am häufigsten werden dabei Aspekte der *Kommunikation*

genannt. Die befragten Lehrkräfte erwarten, dass die im Rahmen der Pandemie etablierten digitalen Kommunikationskanäle längerfristig Bestand haben werden. Dazu zählt zum einen der Austausch zwischen Lehrenden und Lernenden über Messenger-Dienste. „Diese Form der Kommunikation [entspricht] der Alltagswelt der Schülerinnen und Schüler“ (P43) und bietet eine schnelle und unkomplizierte Kontaktmöglichkeit. Zum anderen wird auch die Kommunikation innerhalb des Kollegiums über Messenger oder Videokonferenzen thematisiert, was es beispielsweise ermöglicht „Besprechungen von zu Hause aus durchzuführen“ (P83). Zudem wird auch den erworbenen *digitalen Kompetenzen* eine längerfristige Bedeutung beigemessen, wobei der „Kompetenzzuwachs im Bereich Digitalisierung bei allen am Schulleben Beteiligten“ gesehen wird (P136). Es wird davon ausgegangen, dass sowohl die Lernenden „selbstständiger mit digitalen Medien“ arbeiten (P41) als auch der „Umgang der Lehrkräfte mit den digitalen Medien“ sich verbessert hat und dadurch „Ängste [...] abgebaut“ wurden (P159). Um digitale Möglichkeiten weiterhin nutzen zu können, wird ein weiterer Ausbau der *technischen Ausstattung*, wie beispielsweise die flächendeckende Bereitstellung von Arbeitsendgeräten, erwünscht: „Wenn die Schülerinnen und Schüler im Präsenzunterricht nur ihr Handy (teils ohne WLAN) als Zugriffsmöglichkeit haben, sehe ich keine Option, die schönen, neuen und brauchbaren digitalen Möglichkeiten, die von zu Hause aus gut nutzbar waren, auch im Unterricht zu benutzen“ (P86).

Kritisch wird eine potenziell veränderte *Erwartungshaltung gegenüber Lehrkräften* gesehen. Hier führen einige Befragte den Wunsch nach ständiger Erreichbarkeit an, der u. a. durch die veränderten Kommunikationsbedingungen befördert wird: „Schüler:innen gewöhnen sich an schnelle Antworten [...] und erwarten das womöglich auch zukünftig“ (P297). Schließlich werden vereinzelt allgemeine Aspekte (beispielsweise bezüglich der Bedeutsamkeit der persönlichen Begegnungen zwischen Lehrenden und Lernenden) oder negative Effekte der digitalen Unterrichtspraxis angesprochen. Hier sind exemplarisch Befürchtungen wie „Niedergang der Schriftsprache, Verlust der Schreibfähigkeit, mangelhafter Ausdruck“ (P113) zu nennen. Manche Lehrkräfte gehen dagegen nicht von langfristigen Auswirkungen des Digitalisierungsschubs auf die Unterrichtspraxis aus – beispielsweise aufgrund „bürokratischer Hürden“ (P48) – oder antizipieren gar eine vollständige Rückkehr zur gewohnten Unterrichtspraxis: „Sobald Corona wieder einen Unterricht in Präsenz mit vollen Klassen zulässt, werden sich viele wieder zurücklehnen und die alten Arbeitsblätter rausholen“ (P78).

3.6 Diskussion

Im Rahmen einer Online-Befragung von 305 Lehrkräften in Baden-Württemberg wurde ermittelt, wie sich die digitale Unterrichtspraxis an kaufmännischen beruflichen Schulen zum Zeitpunkt der dritten Welle (April 2021) der Corona-Pandemie gestaltete. Die Befragten berichteten mit Blick auf die technische Ausstattung und organisationale Unterstützung seitens der beruflichen Schulen überwiegend über positive Rahmenbedingungen. Das Vorhandensein eines ausreichend leistungsfähigen Internetzugangs sowie die Unterstützung beim Erwerb der digitalen Fertigkeiten für eine kompetente und reflektierte Mediennutzung werden überwiegend bestätigt. Dieser Befund steht im Einklang mit den Ergebnissen einer Umfrage des Verbands der Lehrerinnen und Lehrer an beruflichen Schulen (BLV) in Baden-Württemberg. Hier zeigte sich, dass sich im Verlauf der Pandemie die technischen Bedingungen verbesserten und die Lernenden zunehmend besser erreicht werden konnten (BLV, 2021).

Als grundsätzlich positiv kann auch der hohe Anteil des synchron durchgeführten Unterrichts gewertet werden. Mit Blick auf die häufig thematisierten Auswirkungen von Schulschließungen auf die Leistungsstände der Schüler:innen (Bujard et al., 2021; Wößmann et al., 2020, 2021) ist dies ein erfreulicher Befund. Die Ergebnisse zu den eingesetzten digitalen Tools im Unterricht zeigen, dass insbesondere Lernplattformen zur Klassenverwaltung und Organisation des Unterrichts, Videokonferenztools sowie Clouds zum Teilen von Dateien eingesetzt wurden. Tools zum Abspielen von Videos und Präsentationsprogramme wurden nur gelegentlich genutzt. Noch seltener zum Einsatz kamen Tools zur gemeinsamen schriftlichen Diskussion, zur Erstellung und Durchführung von Umfragen und Abstimmungen, zur gemeinsamen Bearbeitung von Dateien, zur Erstellung und Durchführung von Quizzen und Tests sowie zum Sammeln und Strukturieren von Ideen. Bei der Bewertung der Befunde ist es wichtig festzuhalten, dass der Einsatz digitaler Tools keine hinreichende Bedingung für gelungenen Unterricht darstellt. Angesichts des hohen Anteils an Tools, die selten oder nie Anwendung finden, könnte man das Ergebnis aber durchaus als einen Hinweis auf das ungenutzte Potenzial der Digitalisierung und entsprechenden Forschungsbedarf lesen.

Weiterführend wurde eine Charakterisierung des Unterrichts mittels des SAMR-Modells versucht. Es zeigt sich, dass Lehrkräfte häufig Tools einsetzten, die der Enhancement-Ebene zugeordnet werden können. Digitale Tools zur Transformation der Unterrichtspraxis (z. B. komplexe Lehr-Lern-Arrangements wie Simulationen oder Learning Analytics) finden sich hingegen selten. Die Ergebnisse passen zu Befunden aus anderen Domänen (Kramer et al., 2019; Schrenk, 2021). Dem Einsatz komplexerer Varianten könnten u. a. fehlende Ressourcen an

Schulen oder fehlende didaktische Konzepte entgegenstehen. Die Auswertung der offenen Antworten gibt Aufschluss über die Erwartungen von Lehrkräften an die Unterrichtspraxis nach der Coronapandemie. Für die postpandemische Unterrichtspraxis gehen die befragten Lehrkräfte insbesondere vom Einsatz digitaler Tools als Ergänzung zum Präsenzunterricht und zur Bereitstellung von Materialien sowie den Bestand von digitalen Kommunikationswegen aus.

Die Ergebnisse der vorliegenden Studie unterliegen zahlreichen Limitationen. Zum einen ist die Möglichkeit der Generalisierbarkeit der Aussagen durch den gewählten Feldzugang und die daraus resultierende Stichprobe nicht gegeben. Frauen sind in der Stichprobe überrepräsentiert. Zudem kann auch mit Blick auf Baden-Württemberg eine Stichprobenverzerrung gegenüber der Gesamtpopulation an Lehrkräften nicht ausgeschlossen werden. Es ist möglich, dass an der Thematik interessierte Lehrkräfte überzufällig häufig an der Befragung teilgenommen haben. Bei den Ergebnissen handelt es sich zudem ausschließlich um Selbstauskünfte der Lehrkräfte. Es liegen somit keine objektiven Daten, beispielsweise zur technischen Ausstattung (welche zudem recht grob abgefragt wurde), vor. Weiterhin erwies sich die Charakterisierung des Unterrichts mittels des SAMR-Modells als schwierig. Eine Überprüfung der Modellstruktur führte nicht zu zufriedenstellenden Ergebnissen. Zudem sei auf die Kritik am SAMR-Modell (Hamilton et al., 2016) verwiesen. Aussagen zur Unterrichtsqualität sind also nicht möglich. Die Befunde zeigen zwar insgesamt, dass ein Großteil der verwendeten digitalen Tools der Enhance-ment-Ebene des SAMR-Modells zuzuordnen sind und damit bezüglich der Transformation des Unterrichts durch den Einsatz von digitalen Tools noch ungenutzte Potenziale bestehen. Daraus lässt sich jedoch keinesfalls auf die Unterrichtsqualität oder den Lernerfolg zurückschließen. Wir können zudem keine Aussage darüber treffen, ob der Einsatz digitaler Tools pandemiebedingt war oder bereits vorher praktiziert wurde. Entsprechend vorsichtig sollten die Ergebnisse interpretiert werden.

Mit Blick auf die Belastung der Befragten konnten leider nur wenige Einflussfaktoren in die Befragung aufgenommen werden. Aspekte wie der wahrgenommene Nutzen der Tools (siehe Bürger et al., 2021) bleiben daher unberücksichtigt. Uns liegen zudem keine longitudinalen Ergebnisse vor, und Informationen zu den Kompetenzen der Lehrkräfte (z. B. auf Basis von TPACK) wären ebenfalls von Interesse. Selbstauskünfte sind hier nur ein schwacher Ersatz. Für zukünftige Studien wäre es wünschenswert, die didaktischen Möglichkeiten digitaler Tools zu berücksichtigen und weiterführende Bezüge zur Unterrichtsqualität (insbesondere zu den Tiefenstrukturen des Unterrichts) herzustellen. Abschließend sei herausgestellt, dass sich die Schulen derzeit in einem tiefgreifenden Transformationsprozess befinden (Harder et al., 2020) und die hier dargestellten Befunde lediglich eine Momentaufnahme darstellen.

3.7 Literaturverzeichnis

- Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, 26(1), 87–122. <https://doi.org/10.1007/s12528-013-9077-3>
- BLV (Verband der Lehrerinnen und Lehrer an beruflichen Schulen in Baden-Württemberg). (2021). *Ergebnisse der BLV-Umfrage zu Fernunterricht*. Berufsschullehrerverband Baden-Württemberg. <https://blv-bw.de/ergebnisse-der-blv-umfrage-zum-fernunterricht-16306/>
- Bujard, M., von den Driesch, E., Ruckdeschel, K., Lass, I., Thönnissen, C., Schumann, A., & Schneider, N. F. (2021). *Belastungen von Kindern, Jugendlichen und Eltern in der Corona-Pandemie* (2/2021; BiB.Bevölkerungs.Studien). Bundesinstitut für Bevölkerungsforschung. <https://doi.org/10.12765/bro-2021-02>
- Bürger, N., Haselmann, S., Baumgart, J., Prinz, G., Girnat, B., Meisert, A., Menthe, J., Schmidt-Thieme, B., & Wecker, C. (2021). Jenseits von Professionswissen: Eine systematische Überblicksarbeit zu einstellungs- und motivationsbezogenen Einflussfaktoren auf die Nutzung digitaler Technologien im Unterricht. *Zeitschrift für Erziehungswissenschaft*, 24(5), 1087–1112. <https://doi.org/10.1007/s11618-021-01050-3>
- Busch, M. (2020). *55 Webtools für den Unterricht: Einfach, konkret, step by step* (4. komplett überarbeitete und aktualisierte Auflage). Auer.
- Chauhan, S. (2017). A meta-analysis of the impact of technology on learning effectiveness of elementary students. *Computers & Education*, 105, 14–30. <https://doi.org/10.1016/j.compedu.2016.11.005>
- Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (Fourth edition). Wiley.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Academic press.
- Dilger, B. (2021). Lernen aus den Erfahrungen mit Emergency-Remote-Lehre aus Studierenden- und Dozierenden-Perspektive. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117(1), 3–13. <https://doi.org/10.25162/zbw-2021-0001>
- Drossel, K., Eickelmann, B., Schaumburg, H., & Labusch, A. (2019). Nutzung digitaler Medien und Prädiktoren aus der Perspektive der Lehrerinnen und Lehrer im internationalen Vergleich. In B. Eickelmann, W. Bos, J. Gerick, F. Goldhammer, H. Schaumburg, K. Schwippert, M. Senkbeil, & J. Vahrenhold (Hrsg.), *ICILS 2018 #Deutschland. Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking* (S. 205–240). Waxmann. <https://doi.org/10.25656/01:18325>
- Drugova, E., Zhuravleva, I., Aiusheeva, M., & Grits, D. (2021). Toward a model of learning innovation integration: TPACK-SAMR based analysis of the introduction of a digital learning environment in three Russian universities. *Education and Information Technologies*, 26(4), 4925–4942. <https://doi.org/10.1007/s10639-021-10514-2>
- Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The Substitution Augmentation Modification Redefinition (SAMR) Model: A Critical Review and Suggestions for its Use. *TechTrends*, 60(5), 433–441. <https://doi.org/10.1007/s11528-016-0091-y>
- Harder, A., Imboden, S., Glassey-Previdoli, D., & Schumann, S. (2020). Schulleitungshandeln in Zeiten der digitalen Transformation: „Business as usual“ oder „Alles ist neu“? *bwp@, Profil 6: Digitale Festschrift für EVELINE WUTTKE*, 18.

- Helm, C., Huber, S., & Loisinger, T. (2021). Was wissen wir über schulische Lehr-Lern-Prozesse im Distanzunterricht während der Corona-Pandemie? – Evidenz aus Deutschland, Österreich und der Schweiz. *Zeitschrift für Erziehungswissenschaft*, 24(2), 237–311. <https://doi.org/10.1007/s11618-021-01000-z>
- Herzig, B. (2007). Medien in komplexen Lernumgebungen. In H.-H. Kremer (Hrsg.), *Paderborner Forschungs- und Entwicklungswerkstatt: Forschungsfragen und -konzepte der beruflichen Bildung* (Bd. 1, S. 157–174). Eusl.
- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897. <https://doi.org/10.1016/j.compedu.2020.103897>
- Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *Educause review*, 27(1), 1–9.
- Howe, F., & Knutzen, S. (2013). *Digitale Medien in der gewerblich-technischen Berufsausbildung. Einsatzmöglichkeiten digitaler Medien in Lern- und Arbeitsaufgaben*. Bundesinstitut für Berufsbildung. https://www.bibb.de/dokumente/pdf/Expertise_Howe_Knutzen.pdf
- Huber, S. G., Günther, P. S., Schneider, N., Helm, C., Schwander, M., Schneider, J., & Pruitt, J. (2020). *COVID-19 und aktuelle Herausforderungen in Schule und Bildung*. Waxmann. <https://doi.org/10.31244/9783830942160>
- Hughes, J., Thomas, R., & Scharber, C. (2006). Assessing technology integration: The RAT-replacement, amplification, and transformation-framework. *Society for Information Technology & Teacher Education International Conference*, 1616–1620.
- Ifenthaler, D., & Delcker, J. (2021). Check-up Distance Learning. Datengestützte Schul- und Unterrichtsentwicklung für berufliche Schulen vor Ort. *Schulverwaltung BW*, 30(11), 296–299.
- Kerres, M., & Kalz, M. (2003). Mediendidaktik in der Lehrerbildung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 21(3), 410–421.
- Kramer, M., Förtsch, C., Aufleger, M., & Neuhaus, B. J. (2019). Der Einsatz digitaler Medien im gymnasialen Biologieunterricht: Eine deskriptive Auswertung einer quantitativen Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 25(1), 131–160. <https://doi.org/10.1007/s40573-019-00096-5>
- Lubega, J., Kajura, M., & Muyinda, P. B. (2014). Adoption of the SAMR Model to Assess ICT Pedagogical Adoption: A Case of Makerere University. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 4(2), 106–115. <https://doi.org/10.7763/IJEEEE.2014.V4.312>
- Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. U.S. Department of Education. www.ed.gov/about/offices/list/opepd/ppss/reports.html
- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record: The Voice of Scholarship in Education*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Moore, M. G., & Kearsley, G. G. (1996). *Distance education: A system view*. Wadsworth.
- Puentedura, R. (2006). *Transformation, Technology, and Education*. <http://hippasus.com/resources/tte/>

- Puentedura, R. (2014). *SAMR, Learning, and Assessment*. <http://www.hippasus.com/rrpweb-log/archives/2014/11/28/SAMRLearningAssessment.pdf>
- Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu*. (Y. Punie, Hrsg.). Publications Office of the European Union. <https://data.europa.eu/doi/10.2760/178382>
- Scheiter, K. (2021). Lernen und Lehren mit digitalen Medien: Eine Standortbestimmung. *Zeitschrift für Erziehungswissenschaft*, 24(5), 1039–1060. <https://doi.org/10.1007/s11618-021-01047-y>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online*, 8(2), 23–74.
- Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R., Abrami, P. C., Wade, C. A., Surkes, M. A., & Lowerison, G. (2009). Technology's effect on achievement in higher education: A Stage I meta-analysis of classroom applications. *Journal of Computing in Higher Education*, 21(2), 95–109. <https://doi.org/10.1007/s12528-009-9021-8>
- Schrenk, R. (2021). Auswirkungen der COVID-19 Pandemie auf die Arbeit mit Moodle – Aktuelles aus Österreichs Schulen. *GW-Unterricht*, 1, 54–60. <https://doi.org/10.1553/gw-unterricht162s54>
- Seifried, J. (2008). Methodische Gestaltung des Unterrichts an kaufmännischen Schulen. *Wirtschaft & Erziehung*, 60(11), 364–370.
- Seifried, J. (2021). Der lange Weg in ein „New Normal“ – Noch ist unklar, was nach der Pandemie bleibt. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117(3), 309. <https://doi.org/10.25162/zbw-2021-0014>
- Seufert, S., Guggemos, J., & Tarantini, E. (2018). Digitale Transformation in Schulen – Kompetenzanforderungen an Lehrpersonen. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 36(2), 175–193. <https://doi.org/10.25656/01:17096>
- Sloane, P. F. E. (2020). Kaufmännische Berufsbildung nach Corona: Überlegungen für die zukünftige Lehrerbildung an beruflichen Schulen und Berufskollegs. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 116(4), 533–543. <https://doi.org/10.25162/zbw-2020-0023>
- Spener, C., & Schumann, S. (2021). Wissensseffekte des ERP-Einsatzes in der kaufmännischen Berufsschule. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117(3), 395–430. <https://doi.org/10.25162/zbw-2021-0018>
- Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What Forty Years of Research Says About the Impact of Technology on Learning: A Second-Order Meta-Analysis and Validation Study. *Review of Educational Research*, 81(1), 4–28. <https://doi.org/10.3102/0034654310393361>
- 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>
- Wacker, A., Unger, V., & Rey, T. (2020). „Sind doch Corona-Ferien, oder nicht?“. Befunde einer Schüler*innenbefragung zum „Fernunterricht“. In D. Fickermann & B. Edelstein (Hrsg.), „Langsam vermissen wir die Schule ...“. *Schule während und nach der Corona-Pandemie.* (Bd. 16, S. 79–94). Waxmann. https://www.pedocs.de/front-door.php?source_opus=20230

- Wahyuni, S., Mujiyanto, J., Rukmini, D., & Fitriati, S. W. (2020). Teachers' Technology Integration Into English Instructions: SAMR Model. *Advances in Social Science, Education and Humanities Research*, 443, 546–550.
- Weinert, F. E. (1999). Die fünf Irrtümer der Schulreformer. Welche Lehrer, welchen Unterricht braucht das Land? *Psychologie Heute*, 26(7), 28–34.
- Wheeler, S. (2012). Distance Learning. In N. M. Seel (Hrsg.), *Encyclopedia of the Sciences of Learning* (S. 1018–1020). Springer. https://doi.org/10.1007/978-1-4419-1428-6_432
- Wößmann, L., Freundl, V., Grewenig, E., Lergetporer, P., Werner, K., & Zierow, L. (2020). Bildung in der Coronakrise: Wie haben die Schulkinder die Zeit der Schulschließungen verbracht, und welche Bildungsmaßnahmen befürworteten die Deutschen? *ifo Schnelldienst*, 73(09), 25–39.
- Wößmann, L., Freundl, V., Grewenig, E., Lergetporer, P., Werner, K., & Zierow, L. (2021). Bildung erneut im Lockdown: Wie verbrachten Schulkinder die Schulschließungen Anfang 2021? *ifo Schnelldienst*, 74(5), 36–52.

4 Looking behind the scenes - Being able to master ERP systems as a goal of vocational education and training

Study 2 was published in June 2024 in ‘Apprenticeships and the digital transition: modernising apprenticeships to meet digital skill needs’ of CEDEFOP and OECD. The publication is available at <https://data.europa.eu/doi/10.2801/074640>. Below it is reproduced with permission from CEDEFOP.⁸

4.1 Relevance of ERP systems for vocational education and training

For some time now, the use of software in the (commercial) workplace has played a significant role in supporting operational business and decision-making⁹. In this context, employees’ skills are now strongly linked to the respective computer-based tasks and workplace roles (Billett, 2021). The results of the latest European Skills and Job Survey (European Centre for the Development of Vocational Training, 2022) illustrate the long-standing digital transformation in the EU. New digital technologies, computer systems, computing devices, or computer programs were used in the workplace by 44% of employees in 2020-21 across Europe and Germany. In addition, an OECD Economics Department working paper shows robust empirical evidence of the link between digitization and productivity in several countries (Gal et al., 2019), with Enterprise Resource Planning (ERP) systems playing an important role.

As ERP systems are now standard in companies, they must be considered in vocational education and training (VET). ERP systems store data in a central database and map business processes holistically across different application levels. The systems access operational information for operational work steps in real-time on a horizontal (functional) level and condense this information to different aggregation levels on a vertical (informational) level (Wigand et al., 2003). Teaching and learning with ERP systems can be linked to simulation-based learning. ERP systems in VET mimic business processes in authentic environments where users interact with typical software and authentic documents to solve complex tasks. In general, simulation-based learning refers to any “instructional tools or devices with which the learner physically interacts to mimic real life”, and “the need to interact with authentic objects” is emphasized

⁸ Note: The chapter headings, figures, and tables of the published version have been numbered according to the numbering of this thesis. Minor changes or deviations from the published paper publication are possible. Spelling and grammar have been adjusted to American English standards. Any misspellings and typographical errors detected in the course of a further thorough proofreading have been corrected.

⁹ This chapter draws on two recently published German papers, namely Mayer (2022) and Mayer & Seifried (2022).

(Cook et al., 2013, p. 876). Empirical evidence shows that simulation-based learning in higher education promotes the acquisition of complex skills (Chernikova et al., 2020).

When it comes to students in VET, ERP systems usage at the workplace for education and training purposes is often limited to routine activities, such as retrieving information, since improper use by trainees could severely disrupt operational processes. The focus here is on developing routines, such as data maintenance (Jasperson et al., 2005). The learning potential of dealing with ERP systems can often not be fully exploited at the company site. Thus, both learning sites, vocational schools and companies, need to complement each other. On the operational side of a business, ERP systems are implemented for task specific purposes, considering industry-specific features and mapping operational specifics (in technical jargon: ‘customizing’). In addition, employees, and especially trainees, are often assigned limited roles that do not sufficiently allow a more comprehensive insight into work processes. Nevertheless, ERP systems can be used in company training to illustrate business processes. In order to do justice to a broader professional logic, ERP systems are used in training at vocational schools to illustrate general business processes, irrespective of company or department specifics. Against this background, more learning opportunities based on the use of ERP, are needed in vocational schools. In the German dual vocational training system, there is a positive experience with intensive work with ERP systems in the vocational school learning environment. Learning with ERP systems at vocational schools is not just about the acquisition of routines (as those preferred in the workplace use of ERP) but also about taking a look ‘behind the scenes’ to promote a deep understanding of the business processes mapped in the ERP system. In schools, mistakes and errors are allowed (and are viewed as a central component of learning processes; Metcalfe, 2017), and processes can be decelerated comprehensibly. Trainees can better understand cross-departmental business processes through ERP systems (Pongratz et al., 2009). Additionally, the employability of trainees can be promoted (Zutavern & Seifried, 2022).

Due to the potential of ERP systems as a teaching and learning subject, it is essential to qualify (prospective) teachers and in-company trainers to implement these systems in a didactically meaningful way (see Box 1).

Box 1. Why ERP systems have not fully arrived in schools yet?

« One reason for the scarce integration of software into lessons may be that teachers still lack of qualification themselves. »

Source: (Knigge et al., 2017, p. 492).

In this paper, the potential of teaching digital work processes to trainees using ERP systems in vocational schools is discussed. We do this against the background of the situation in the German VET system, considering (1) curricular requirements, (2) didactic possibilities and challenges, as well as (3) the necessity of an adequate competence assessment.

We will briefly introduce the theoretical underpinnings of teaching and learning with ERP systems. We will then summarize our research on teacher empowerment based on previous work (Mayer, 2022; Mayer & Seifried, 2022) and discuss limitations and implications for vocational education, training, and educational policymakers.

4.2 Theoretical Foundation for Teaching and Learning with ERP systems

Curriculum-Instruction-Assessment (CIA) triad

To discuss teaching and learning with ERP systems, we rely on the Curriculum-Instruction-Assessment (CIA) triad (see Fig. 4-1). We use this approach to address the following:

- (1) The knowledge and skills that need to be learned (curriculum perspective).
- (2) The learning activities that guide and promote the acquisition of the relevant knowledge and skills (instructional perspective).
- (3) The appropriate measurement (assessment perspective) of the acquired knowledge and skills (Achtenhagen, 2012; Pellegrino, 2006).

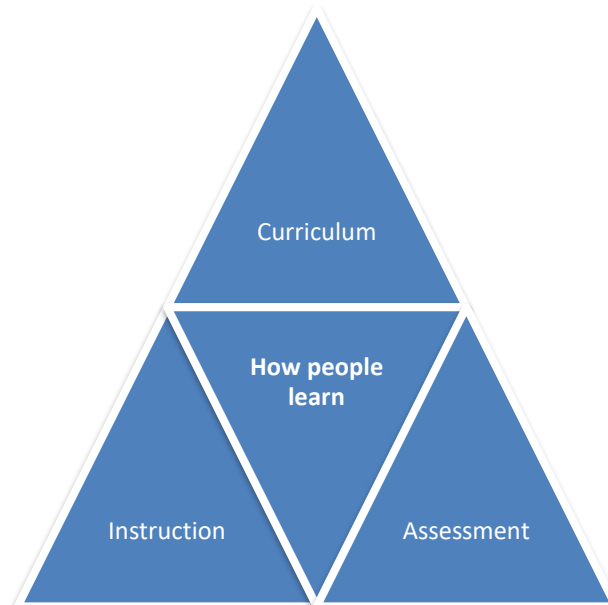


Figure 4-1: Curriculum-Instruction-Assessment (CIA) triad in the style of Pellegrino (Achtenhagen, 2012; 2002, 2006).

Curriculum

Training occupations and therefore training regulations have to be revised on a regular basis to meet the needs of the workplace. The changes in an increasingly digitalized world of work are being perceived accordingly at the institutional level and are also increasingly anchored in VET curricula. For example, the ‘guiding questions’ of the Federal Institute for Vocational Education and Training’s (BIBB) thematic cluster on digitization refer to the examination of the potential influence of new technologies and the resulting new design options for education and training systems (Bundesinstitut für Berufsbildung, 2022; Zinke, 2019). Consequently, one of the four recently introduced cross-occupational standards for vocational training (new themes across occupational programs) of the BIBB focuses on a ‘digitalized world of work’ which not only has a recommendatory character for all existing training regulations but is mandatory in all new or updated training regulations that come into force as of August 1, 2021.

Within the framework curricula of the Standing Conference (a consortium of policy makers responsible for education and schooling, among other things) of the Ministers of Education and Cultural Affairs of the federal states in the Federal Republic of Germany (Kulturministerkonferenz / KMK), the use of integrated company software is explicitly listed as content to be taught in some training occupations and learning fields.

From the perspective of vocational schools offering programs in Commerce (e.g., in retail, banking insurance etc., hereafter referred to as ‘commercial trainees), examples of this can be found in IT-related training occupations such as an apprenticeship in the field of IT systems management or digitalization management where vocational learning is often structured in so called learning fields/training modules (as an example see learning field 12: Carrying out and monitoring sales processes: “You document the entire sales process with the support of software (integrated company software)” (Kultusministerkonferenz, 2019b, p. 12). Similarly, other commercial apprenticeships such as wholesale and foreign trade clerks have been considering the use of integrated company software. In learning field 2, titled “Processing orders in a customer-oriented manner,” trainees create and check invoices and delivery notes while explaining the benefits of integrated business software (Kultusministerkonferenz, 2019a, p. 11).

Instruction

In the context of business process didactics, functionally operationalized work processes within ERP systems are connected to corresponding domain-specific knowledge based on authentic situations (the situation principle) within a simulated environment (Wilbers, 2014). From a ped-

agogical perspective, it is essential to make a clear distinction between business process didactics and pure user training. Classic user training, often referred to as ‘click training’, is insufficient for acquiring comprehensive action competence (Weinert, 2001). Contextual knowledge is also needed for trainees to grasp the corresponding business process. Accordingly, when working through the first steps of a sales process (see Fig. 4-2), not only inquiries, quotations, and incoming orders are ‘worked through’ in the ERP system, but in this context, equally subject-didactically relevant contents are taught. For example, while working through the sales process in the ERP system, the agreement of two declarations of intent as the basis of a sales contract can be explained or disruptions to the sales contract can be discussed with consideration of the relevant legal texts. This could help to link procedural knowledge with contextual and factual knowledge.



Figure 4-2: Sales process (simplified) in the style of SAP4school material (SAP4school IUS, 2023).

The possible uses of ERP systems are not limited to processing tasks in the computer room of schools but allow for some instructional variation possibilities. Asynchronous and synchronous formats are possible. Additionally, and more precisely, Wilbers (2009) differentiates between (1) illustrative use, (2) sequential use, (3) use in learning companies, and (4) use in projects/cases/learning situations. Illustrative use is a low-cost and low-threshold variant of presenting ERP systems using screenshots. The sequential application option requires a theoretical concept phase, followed by a small-stepwise task processing in the ERP system. The most established implementation of ERP systems in learning and teaching at schools in Germany is the use of ERP systems in learning companies (practice firms and junior companies). Teaching and learning in this style goes hand in hand with the complex teaching-learning arrangement of (fictitious/real) product, money flows, and (real) suppliers and customers. On the other hand, the use of ERP systems in projects, cases and learning situations takes place in addition to teaching and is more likely to be found in the university context. Adapted case studies and more complex cases are worked on (Wilbers, 2009).

Assessment

According to Pellegrino (2006), assessment should measure what students are expected to learn according to the curriculum. However, performance-based assessments of skill acquisition are still rare in relation to the use of ERP. Most often, a particular learning module must be worked through to see if specific knowledge and skills have been learned. However, performance-based measurement based on previously achieved tasks is a prospect for ERP assessments. The challenges of pedagogical, technical, and individual teacher and student constraints remain a barrier and must be addressed.

TPACK

The Technological Pedagogical Content Knowledge model (TPACK) encompasses different dimensions of teachers' professional knowledge (Koehler & Mishra, 2009; Mishra, 2019; Mishra & Koehler, 2006). It extends a model for describing teachers' professional knowledge by Shulman (1986) by adding a technological component. The TPACK model distinguishes further between content knowledge (CK), pedagogical knowledge (PK), and technological knowledge (TK), as well as the intersections of these types of knowledge (see Fig. 4-3).

Technological knowledge (TK) includes knowledge about media and technological applications and underlying technological concepts (Mishra & Koehler, 2006). Content knowledge (CK) refers to the knowledge of the respective professions, often certified through government degrees (Shulman, 1986). Pedagogical knowledge (PK), on the other hand, describes knowledge about teaching and learning, taking into account pedagogical-psychological and didactical approaches such as principles of classroom management or the organization of lessons (Shulman, 1986).

Pedagogical content knowledge (PCK) combines pedagogical-psychological knowledge and subject knowledge and refers to sub-areas of the underlying professional domains (Schmid, Krannich, et al., 2020). Technological content knowledge (TCK), on the other hand, represents the combination of technological and subject knowledge and refers to possibilities of teaching and learning with the inclusion of technology (Schmid, Krannich, et al., 2020). Technological pedagogical knowledge (TPK) refers to media-didactic knowledge (Schmid, Krannich, et al., 2020).

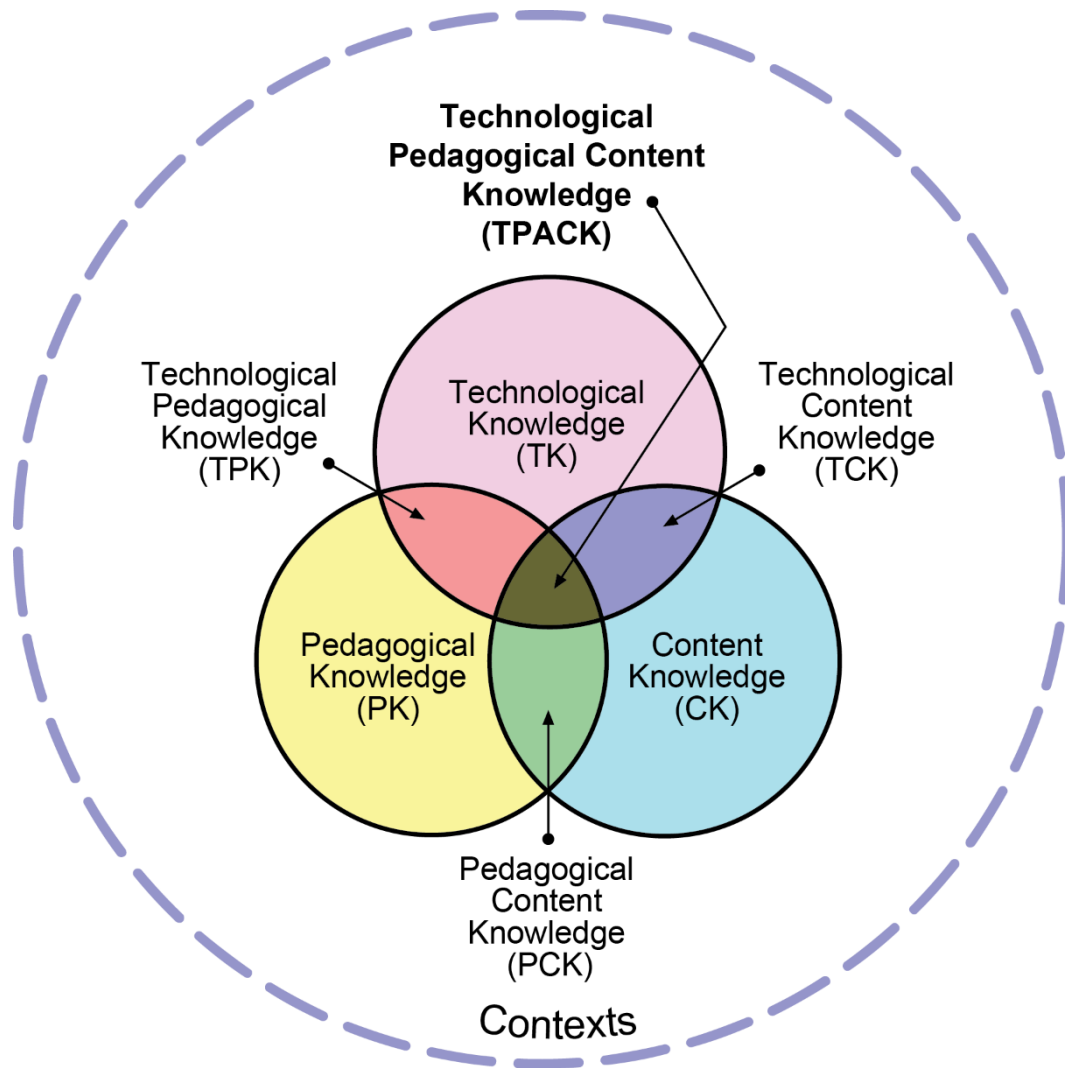


Figure 4-3: Technological Pedagogical Content Knowledge (Mishra & Koehler, 2006). Source: available at <http://tpack.org> (TPACK.org, 2023).¹⁰

The core of the model, however, is the technology-related subject didactic knowledge (TPCK), which combines all the previously mentioned areas of knowledge (Schmid, Krannich, et al., 2020). In addition, the contextual framework itself (Contexts) is also increasingly addressed within the TPACK model, in order to also be able to take into account the (technological) conditions of schools, teachers and students (Mishra, 2019).

In the context of ERP systems, for example, TK refers to the knowledge of the structure and functionality of ERP systems and relational databases. The CK component refers to the knowledge surrounding using ERP systems to accomplish professional tasks. PK includes, among other things, the possibilities of using ERP for instruction conducive to learning and for the practical assessment of learning outcomes. From a PCK perspective, teaching is closely

¹⁰ Reproduced by permission of the publisher, © 2012 by tpack.org

linked to knowledge about typical errors and misconceptions of learners and hurdles during learning and skill acquisition. In the case of using ERP, these include, for example, errors in raising an invoice or difficulties in understanding the accounting scheme in the general ledger and sub-ledger of accounts receivable. TCK includes, for example, knowledge about using ERP systems to promote domain-specific knowledge, such as the modelling and handling of holistic business processes. While TPK, among other aspects, also involves the potential of new visualization and presentation possibilities using ERP systems. Finally, TPCK reflects the didactically meaningful use of ERP systems across all previously mentioned sub-areas for teaching specific subject didactic content. Additionally, in the context of ERP systems, attention should be drawn to the availability of time, human and material resources (especially software and hardware).

4.3 Methodology

The paper draws evidence from teaching prospective teachers and trainers in VET on the possibilities of using ERP systems, as part of their Masters program in business education offered at the University of Mannheim/Germany. The majority of the program graduates are employed either as teachers of Business in trainees programs in Commerce (wholesale, industry, trade, retail, bank and insurance, and so forth) or as training managers or HR staff in companies participating in VET.

We designed a course related to teaching and learning with ERP Systems that deals with the didactic examination of the possibilities of using a specific ERP system in the classroom or instruction in business practice at the workplace.

In the autumn semester of 2021 and spring semester of 2022, 26 students successfully participated in the seminar. On average, the participants were 25.6 years old and in their second Master's semester. Almost all participants had little to no experience with ERP systems or SAP ERP.

A voluntary pseudonymized entry and exit survey accompanied the seminar. Of the 26 students who took the elective, 20 participants completed both the entry and exit surveys. In addition, data from an anonymized student teaching evaluation is available, and conducted towards the end of the semester. To capture the TPACK dimensions, the questionnaire of Schmid, Brianza, and Petko (2020) is used, which comprises 28 items (4 items per TPACK dimension). An example item of the PK dimension is: *I can adapt my teaching to what learners currently understand and what they do not understand*. The Likert scale ranges from 1 (strongly disagree) to 5 (strongly agree). The reliabilities of the scales can be described as predominantly satisfactory

to good (Cronbach's $\alpha = .67$ to $.90$). In addition, a structured reflection diary was used. The students reflected on their self-experienced difficulties during processing.

For the data analysis, the free statistical software R was used in the R Studio environment with extending packages, particularly "Tidyverse" (R Core Team, 2021; R Studio Team, 2021; Wickham et al., 2019). We resort to non-parametric test procedures for analyzing dependent samples, as the prerequisites for applying parametric test procedures are not given in the present case (Döring & Bortz, 2016). The test for normal distribution using Shapiro-Wilk tests shows significant deviations from the standard normal distribution for the data. Therefore, the non-parametric Wilcoxon test (one-sided hypothesis test, i.e., the values of the initial survey should lie above those of the initial survey) was used. Further, the effect sizes were calculated as r -values to classify the results.

4.4 Findings

Insights into the results of the teaching evaluation, in which 15 of the 20 participants took part, show that the students rate the seminar quite well (value range from 1 = fully agree to 5 = disagree at all): *Overall, I am very satisfied with the lecturer* ($M = 1.1$). *Overall, I am very satisfied with the course* ($M = 1.5$). *I have learned a lot in this course* ($M = 1.6$).

A holistic analysis of the self-assessment of one's own skills in dealing with ERP systems (range 1 = no skills at all [novice] to 10 = profound knowledge [expert]) yields the following result: There is a significant increase over the course of the lecture ($M_0 = 2.2$ vs $M_1 = 6.5$, $V = 210$, $p < .001$, $n = 20$). The effect strength is $r = .88$ and corresponds to a strong effect (according to Cohen, 1992).

To analyze the TPACK dimensions, a mean value was calculated for each subscale, and a pre-post test was conducted. Overall, only the TK dimension showed a significant effect ($V = 114.5$, $p < .05$). The effect size is $r = .41$ (moderate effect). All the other dimensions show no significant changes.

In order to account for biases in the self-assessment (over- or underestimation), participants were assigned to two groups in a further analysis step. For this purpose, the mean value of the initial survey per participant was formed over all TPACK items and assigned to one of two groups utilizing a median split ($Mdn = 3.75$). The students in group 1 ($n = 11$) tended to have lower self-assessment values, while those in the second group ($n = 9$) had higher values for self-assessment. The analysis of the development of competencies in the course at group levels shows that especially the students with lower values at the beginning benefit from the course.

Significant increases were recorded for the TK dimension ($V = 51, p < .01$) and TCK dimension ($V = 44.5, p < .05$) (with high effect sizes in each case: TK: $r = .74$; TCK: $r = .51$). For the remaining dimensions, no significant changes can be detected. No significant results were found when comparing the groups (see Mayer & Seifried, 2022 for further details). Here, one could assume that an initial overestimation of one's abilities was relativized during the course (further information on this aspect can be found in the conclusion section).

The brief insights into the reflection diaries put this result into perspective. These indicate that the students often encountered problems when working on the tasks. It becomes clear that they may have overestimated their knowledge and skills at the beginning of the course. The majority of the processing problems encountered, on the other hand, can be traced back to the procedure in the ERP system. Typical procedural processing problems can be further subdivided as follows:

- 1) Processing problems in the ERP system due to a lack of essential contents and concepts: "There is a lack of background knowledge. Deeper understanding of taxes, for example. When is it full taxation or reduced? Or how are list prices calculated. [...]" (problem number / PNr. 116).
- 2) Processing problems in the ERP system due to a lack of knowledge of the general functioning and structures of ERP systems, for example: "[...] It was unclear to me that I have to create the offer with reference, I overlooked that or did not think about it. I knew that it was a quotation corresponding to the previously entered request, but I did not know that I could reference it in the system. In the solution notes, it's easy to miss the reference to it because it's right under the process steps." (PNr. 031). The referencing of data is an essential feature. The relational database forms the basis of ERP systems, and the underlying concept should be clear. Referencing not only facilitates the preparation of quotations (among other receipts) by taking over data that has already been stored, but also key indicators for sales control can only be reliably determined if they are referenced accordingly in the system. Similarly, an offer does not necessarily have to result from an inquiry, so creating an offer with and without a reference is possible (see also Mayer, 2022).
- 3) Processing problems in the ERP system due to a lack of skills about the system-specific application of transactions. This includes general skills lacking in using the software: "In the delivery task, we were supposed to create a substitute document in the logistics category, a single document related to the sales order. When I entered the navigation device, I had to look for it again because I couldn't remember the numbers. When I

entered it, I didn't notice the asterisks of Navi* and the corresponding *number, so nothing was shown to me at first" (PNr. 050). In this case, as is common in computer applications, * symbols are used as placeholders (jokers or wildcards) for the search. For learners, this is convenient for searching, while teachers can use wildcards in transactions to check an entire class's progress without monitoring individual student data.

Nevertheless, more specific skill lacks in dealing with the ERP system also cause problems in processing the tasks, as the following problem descriptions show: It is reported that "the problem [was] that the material created was saved and therefore some views could not be changed" (PNr. 061). Here, system-side specifics show up, which can only be acquired through appropriate routines. The latter problem could be solved easily by changing the transaction 'create' to 'change' and was solved in the present case by creating another master data entry for the material.

In the end, careless mistakes also prove to be problems in processing the tasks, as they will undoubtedly occur just as frequently in the school context: "I registered too quickly and without thinking. So my set language was English" (PNr. 018).

4.5 Conclusions

The paper referred to the benefits of introducing ERP in teaching trainees, first at vocational schools, but also at the workplace. To ensure a high quality ERP education, (prospective) teachers and corporate trainers need to know the implications of the curriculum, instruction, and assessment dimensions of learning (CIA triad) and how they relate to ERP systems. Therefore, we have designed a master seminar for prospective company trainers and vocational school teachers to promote the teaching and learning with ERP systems among our students.

In this paper we reported the findings of a teaching-learning concept for using ERP systems among business education students and analyzed the self-assessed knowledge acquisition along the TPACK dimensions in a pre-post design (Mishra & Koehler, 2006). Basically, the students reported high satisfaction with the course offered for the first time at the University. Positive effects can also be seen concerning the self-assessed change in ERP skills. Regarding the TPACK dimensions, however, the picture is less substantial — only a few significant changes can be detected. Exemplary insights into the students' reflections recorded in a reflection diary show, however, that a misjudgment of their competencies in the sense of an overestimation at the beginning of the semester could be at least partly responsible for the lack of more significant effects. The reflection diaries point to clear gaps in the knowledge of some students. An analysis carried out against the background of these considerations, taking into account the different

self-assessment levels, then further shows those participants with a tendency towards lower self-assessment in particular benefit from the course. This subgroup shows strong effects on the self-assessed knowledge acquisition regarding technological and technological-subject knowledge. However, no significant changes can be found for participants with higher self-assessments. This indicates an adjustment of the competence assessment. If a misjudgment of prospective teachers' and in-company trainers competencies can be corrected within the framework of student reflection, this would have a desirable side effect from a pedagogical point of view. This is especially true for prospective teachers and in-company instructors, as those who receive specialized pedagogical training in teaching and training with ERP systems will benefit more when assessing the skills of trainees compared to teachers and in-company trainers without such a training.

Additionally, typical challenges in processing authentic commercial tasks in the ERP system were qualitatively evaluated. These are of great importance for designing corresponding lesson plans for vocational education in vocational schools and in-company trainings with ERP systems since the typical self-reflective problems should be considered in the further design of learning processes. Typical problems in processing can be traced back to a lack of basic subject knowledge of prospective teachers, a lack of knowledge about the general functioning and structure of ERP systems, and a lack of knowledge about the application of processes in the underlying ERP system. In addition, the reflection diaries contain references to aspects which, although they can be described as careless mistakes, are also likely to occur in this way in everyday school life and should be avoided as much as possible in the daily work in a value chain of a company.

Given the specific setting and small sample size, the results presented here should be interpreted cautiously. Accordingly, given self-report limitations (especially concerning social desirability bias), future research should be supplemented by an authentic competence assessment in subsequent studies within authentic VET environments (e.g. assessment of trainees skills in in-company trainings and vocational schools).

For (prospective) teachers and in-company trainers, assessing their own knowledge and skills is a necessary condition for competence development in their professional biography (Ross & Bruce, 2007). We argue that there is an urgent need to educate prospective teachers and in-company trainers further and consider the use of ERP in current teacher and trainer programs in VET to address this problem. For example, integrating ERP systems into the training phase at the university level might solve the problem of prospective teachers lacking workplace skills

relevant to technology use and might help to educate the next generation of trainees appropriately. However, further research is needed to assess the skill acquisition of (prospective) teachers (university / learning-site school) and trainers (learning-site company) and trainees for teaching and learning with ERP systems (for deeper insights see Mayer, 2022; and Mayer & Seifried, 2022).

4.6 References

- Achtenhagen, F. (2012). The curriculum-instruction-assessment triad. *Empirical Research in Vocational Education and Training*, 4(1), 5–25.
- Billett, S. (2021). Mediating worklife learning and the digitalisation of work. *British Journal of Educational Technology*, 52(4), 1580–1593. <https://doi.org/10.1111/bjet.13115>
- Bundesinstitut für Berufsbildung. (2022, August 17). *Digitale Transformationen [Digital transformations]*. <https://www.bibb.de/de/120255.php>
- Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-Based Learning in Higher Education: A Meta-Analysis. *Review of Educational Research*, 90(4), 499–541. <https://doi.org/10.3102/0034654320933544>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., & Hatala, R. (2013). Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis. *Medical Teacher*, 35(1), e867–e898. <https://doi.org/10.3109/0142159X.2012.714886>
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften [Research methods and evaluation in the social and human sciences]* (5. vollständig überarbeitete, aktualisierte und erweiterte Auflage). Springer.
- European Centre for the Development of Vocational Training. (2022). *Setting Europe on course for a human digital transition: New evidence from Cedefop's second European skills and jobs survey*. Publications Office. <https://data.europa.eu/doi/10.2801/253954>
- Gal, P., Nicoletti, G., Renault, T., Sorbe, S., & Timiliotis, C. (2019). *Digitalisation and productivity: In search of the holy grail – Firm-level empirical evidence from EU countries* (OECD Economics Department Working Papers 1533). OECD Publishing. <https://doi.org/10.1787/5080f4b6-en>
- Jasperson, J., Carter, P., & Zmud, R. (2005). A Comprehensive Conceptualization of Post-Adoptive Behaviors Associated with Information Technology Enabled Work Systems. *MIS Quarterly*, 29(3), 525. <https://doi.org/10.2307/25148694>
- Knigge, M., Prifti, L., Kienegger, H., & Krcmar, H. (2017). Teaching enterprise organization and enterprise resource planning systems in schools: Playing a serious game with pupils. *2017 IEEE Global Engineering Education Conference (EDUCON)*, 486–495. <https://doi.org/10.1109/EDUCON.2017.7942891>
- Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60–70.
- Kultusministerkonferenz. (2019a). *Rahmenlehrplan für den Ausbildungsberuf Kaufmann für Groß- und Außenhandelsmanagement und Kauffrau für Groß- und Außenhandelsmanagement [Framework curriculum for the training occupation of management assistant*

- in wholesale and foreign trade management*]. https://www.kmk.org/themen/berufliche-schulen/duale-berufsausbildung/downloadbereich-rahmenlehrpläne.html?type=150&tx_fedownloads_pi1%5Bdownload%5D=45053&tx_fedownloads_pi1%5Baction%5D=forceDownload&tx_fedownloads_pi1%5Bcontroller%5D=Downloads&cHash=c5732d9a98f6f4837cf89b90b84ecc7f
- Kultusministerkonferenz. (2019b). *Rahmenlehrplan für die Ausbildungsberufe Kaufmann für IT-System-Management und Kauffrau für IT-System-Management Kaufmann für Digitalisierungsmanagement und Kauffrau für Digitalisierungsmanagement [Framework curriculum for the training occupations of IT systems management clerk and IT systems management clerk Digitization management clerk]*. https://www.kmk.org/themen/berufliche-schulen/duale-berufsausbildung/downloadbereich-rahmenlehrpläne.html?type=150&tx_fedownloads_pi1%5Bdownload%5D=45053&tx_fedownloads_pi1%5Baction%5D=forceDownload&tx_fedownloads_pi1%5Bcontroller%5D=Downloads&cHash=c5732d9a98f6f4837cf89b90b84ecc7f
- Mayer, C. W. (2022). Lehren und Lernen mit Enterprise Resource Planning (ERP) Systemen—Typische Bearbeitungsprobleme als Grundlage der Lernprozessgestaltung [Teaching and learning with Enterprise Resource Planning (ERP) systems Typical processing problems as a basis for designing the learning process]. *Bwp@ Berufs- Und Wirtschaftspädagogik – Online*, 43, 1–17.
- Mayer, C. W., & Seifried, J. (2022). Lehren und Lernen mit ERP-Systemen – Befunde der Evaluation einer Lehrveranstaltung zur Förderung der Technologiekompetenz von Studierenden der Wirtschaftspädagogik [Teaching and learning with ERP systems—Findings from the evaluation of a course to promote the technological competence of business education students]. In U. Mathis, N. Ondrusch, D. Kilian, H. Krcmar, K. Turowski, & S. Weidner (Eds.), *Proceedings of the SAP Academic Community Conference 2022 DACH* (pp. 10–24). Technische Universität München. <https://mediatum.ub.tum.de/doc/1685828/document.pdf#page=10>
- Metcalf, J. (2017). Learning from Errors. *Annual Review of Psychology*, 68(1), 465–489. <https://doi.org/10.1146/annurev-psych-010416-044022>
- Mishra, P. (2019). Considering Contextual Knowledge: The TPACK Diagram Gets an Upgrade. *Journal of Digital Learning in Teacher Education*, 35(2), 76–78. <https://doi.org/10.1080/21532974.2019.1588611>
- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record: The Voice of Scholarship in Education*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Pellegrino, J. W. (2002). Knowing what students know. *Issues in Science and Technology*, 19(2), 48–52.
- Pellegrino, J. W. (2006). Rethinking and Redesigning Curriculum, Instruction and Assessment: What Contemporary Research and Theory Suggests. *Commission on the Skills of the American Workforce*, Chicago, 1–15.
- Pongratz, H., Tramm, P. T., & Wilbers, K. (Eds.). (2009). *Prozessorientierte Wirtschaftsdidaktik und Einsatz von ERP-Systemen im kaufmännischen Unterricht [Process-oriented business didactics and the use of ERP systems in commercial teaching]*. Shaker.
- R Core Team. (2021). *R: A language and environment for statistical computing*. [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- R Studio Team. (2021). *RStudio: Integrated Development Environment for R*. [Computer software]. RStudio, PBC. <http://www.rstudio.com/>

- Ross, J. A., & Bruce, C. D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education*, 23(2), 146–159. <https://doi.org/10.1016/j.tate.2006.04.035>
- SAP4school IUS. (2023). <https://s4s.integrus.de/>
- Schmid, M., Brianza, E., & Petko, D. (2020). Developing a short assessment instrument for Technological Pedagogical Content Knowledge (TPACK.xs) and comparing the factor structure of an integrative and a transformative model. *Computers & Education*, 157, 103967. <https://doi.org/10.1016/j.compedu.2020.103967>
- Schmid, M., Krannich, M., & Petko, D. (2020). Technological Pedagogical Content Knowledge. Entwicklungen und Implikationen [Technological Pedagogical Content Knowledge. Developments and implications]. *Journal für LehrerInnenbildung*, 20(1), 116–124. <https://doi.org/10.25656/01:19396>
- Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- TPACK.org. (2023). *TPACK.ORG*. <http://tpack.org/>
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Hogrefe.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wigand, R. T., Mertens, P., Bodendorf, F., König, W., Picot, A., & Schumann, M. (2003). *Introduction to business information systems*. Springer.
- Wilbers, K. (2009). Integrierte Unternehmenssoftware (ERP-Systeme) im kaufmännischen Unterricht [Integrated business software (ERP systems) in commercial teaching]. In H. Pongratz, P. T. Tramm, & K. Wilbers (Eds.), *Prozessorientierte Wirtschaftsdidaktik und Einsatz von ERP-Systemen im kaufmännischen Unterricht* (Vol. 4, pp. 61–76). Shaker.
- Wilbers, K. (2014). *Wirtschaftsunterricht gestalten [Designing business education]*. epubli. https://www.pedocs.de/frontdoor.php?source_opus=8420
- Zinke, G. (2019). *Berufsbildung 4.0 - Fachkräftequalifikationen und Kompetenzen für die digitalisierte Arbeit von morgen: Branchen- und Berufscreening: vergleichende Gesamtstudie [Vocational education and training 4.0 - Skilled qualifications and competences for the digitalized work of tomorrow: Sector and occupation screening: comparative overall study]*. Bundesinstitut für Berufsbildung.
- Zutavern, S., & Seifried, J. (2022). Vocational Education and Training in Germany: Benefits and Drawbacks of the Dual Approach as Preparation for Professional Employment. In C. Harteis, D. Gijbels, & E. Kyndt (Eds.), *Research Approaches on Workplace Learning* (Vol. 31, pp. 347–365). Springer. https://doi.org/10.1007/978-3-030-89582-2_16

5 Analyzing domain-specific problem-solving processes within authentic computer-based learning and training environments by using eye-tracking: A scoping review

Study 3 was published in April 2023 in ‘Empirical Research in Vocational Education and Training’. The publication is available at <https://doi.org/10.1186/s40461-023-00140-2>. Reproduced with permission from Springer Nature.¹¹

5.1 Abstract

Recently, many studies have been published on the use of eye-tracking to analyse complex problem-solving processes within authentic computer-based learning and training environments. This scoping review aims to provide a systematic report of the current state-of-the-art for related papers. Specifically, this work offers a scoping review of studies that analyse problem-solving processes by using eye-tracking (alongside additional process data such as log files, think aloud, facial expression recognition algorithms, or psychophysiological measures) within authentic technology-based learning and training environments for professional and vocational education and training (VET). A total of 12 studies were identified. The most commonly calculated measures in eye-tracking research are position measures, and these are almost exclusively position duration measures such as the proportion of fixation times or total dwell times. Count measures are also mostly related to the number or proportion of fixations and dwells. Movement measures are rarely computed and usually refer to saccade directions or a scan path. Also, latency and distance measures are almost never calculated. Eye-tracking data is most often analysed for group comparisons between experts vs. novices or high vs. low-performing groups by using common statistical methods such as t-test, (M)ANOVA, or non-parametric Mann-Whitney-U. Visual attention patterns in problem-solving are examined with heat map analyses, lag sequential analyses, and clustering. Recently, linear mixed-effects models have been applied to account for between and within-subjects differences. Also, post-hoc performance predictions are being developed for future integration into multimodal learning analytics. In most cases, self-reporting is used as an additional measurement for data triangulation. In addition to eye-tracking, log files and facial expression recognition algorithms are also used. Few studies use shimmer devices to detect electrodermal activity or practice concurrent thinking aloud. Overall, Haider and Frensch’s (1996, 1999) “information reduction hypothesis” is

¹¹ Note: The chapter headings, figures, and tables of the published version have been numbered according to the numbering of this thesis. Minor changes or deviations from the published paper publication are possible. Spelling and grammar have been adjusted to American English standards. Any misspellings and typographical errors detected in the course of a further thorough proofreading have been corrected.

supported by many studies in the sample. High performers showed a higher visual accuracy, and visual attention was more focused on relevant areas, as seen by fewer fixation counts and higher fixation duration. Low performers showed significantly fewer fixation durations or substantially longer fixation durations and less selective visual attention. Performance is related to prior knowledge and differences in cognitive load. Eye-tracking, (in combination with other data sources) may be a valid method for further research on problem-solving processes in computer-based simulations, may help identify different patterns of problem-solving processes between performance groups, and may hold additional potential for individual learning support.

Keywords: Complex Problem-Solving, Computer-Based Learning Environments, Computer-Based Simulations, Vocational Education and Training, VET, Online Measurements, Eye-Tracking

5.2 Introduction

In educational research, collecting behavioural data is becoming increasingly important to learn more about cognitive and metacognitive processes during learning and instruction. Eye-tracking, a method for analysing the course of gaze, is increasingly used in educational research to improve the instructional design in computer-based learning environments and multimedia learning, to understand and promote the development of expertise, and to visualize the eye movements of experts (Jarodzka et al., 2017). Systematic reviews and meta-analyses of eye-tracking studies are available for various domains (e.g., medical education Ashraf et al., 2018; mathematics: Strohmaier et al., 2020). Similarly, reviews on multimedia learning and instructional design were conducted (e.g., Alemdag & Cagiltay, 2018: multimedia learning; Yang et al., 2018: instructional design of e-learning). However, little research has been done on vocational education and training (VET). This is especially true when more complex vocational tasks are the focus. Therefore, the paper at hand provides an overview of studies that have analysed domain-specific problem-solving processes by using eye-tracking (next to additional online data such as log files or psychophysiological measures) within authentic computer-based learning and training environments in professional training or vocational education and training. The review of the current state of research is conducted as a scoping review. Scoping reviews (Arksey & O'Malley, 2005) are considered a useful approach to examining the design and conduct of research on a particular topic (for the key features of the scoping review approach see Munn et al., 2018). Similar to systematic reviews, scoping reviews are transparent and replicable by following a rigorous study search and selection process. Because fewer studies were identified that followed the same study design, we focused on a scoping review.

While systematic reviews aim to answer a specific question, a scoping review identifies, reports, and discusses a broader perspective on a given topic (Arksey & O'Malley, 2005; Munn et al., 2018; Van Ostaeyen et al., 2022), such as analysing domain-specific problem-solving in computer-based simulations by using eye-tracking.

Computer-based learning environments (CBLE) refer to a broad range of technologies to support learning and instruction (Lajoie & Naismith, 2012). This review includes studies that used eye-tracking as the primary method for collecting behavioural data in computer simulations and serious games. To control for different levels of immersion and interactivity, we excluded close-to-reality simulations, such as in situ studies and realistic simulator training (e.g., flight simulation including a full cockpit, virtual reality welding simulation, nursing practice simulation with mannequins or actors) and similarly, we excluded studies that examined cross-sectoral problem-solving testing (e.g., Raven Matrices, Tower of Hanoi). Also, we excluded studies on collaborative problem-solving, as these studies often focus on the phenomena of joint visual attention. This rigorous procedure resulted in a sample of studies, all of which analysed domain-specific problem-solving in computer-based simulations (see Figure 5-1).

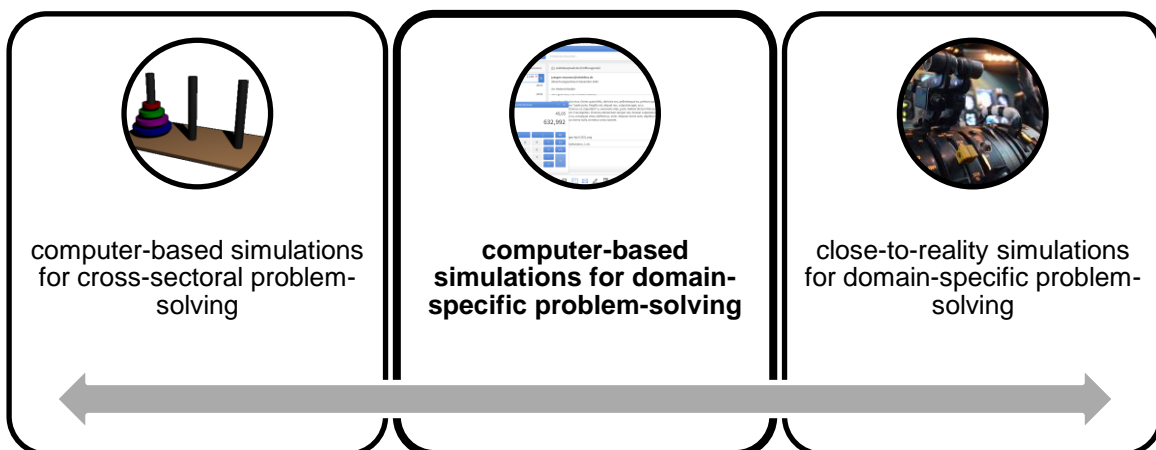


Figure 5-1: The scope of this review (in bold) is illustrated within a continuum between cross-sectoral and domain-specific problem-solving simulations (pictures CC0-licensed).

The scoping review at hand aims to provide a systematic report of the current state-of-the-art in an emerging research field. It contains a variety of eye-tracking and process data measures in a broad range of domain-specific problem-solving tasks. Eye movements can be analysed with various measurements and should be carefully collected and interpreted (Holmqvist et al., 2011). In multimedia learning research, there is a wide range of empirical research methods. Online process tracking techniques such as eye tracking can be combined with other common measurement methods to draw better inferences (Jarodzka, 2021). Thus, we address the following questions:

RQ1: Which eye-tracking measures and additional behavioural measurements were used and how were they analysed?

RQ2: What are the main findings of online data measures in relation to solving complex problems in computer-based simulations in VET?

The paper is structured in five sections. First, a brief overview of the theoretical background on problem-solving, computer-based simulations, and eye-tracking is given. Second, the methodological approach of this review (identification, screening, eligibility, and inclusion) is presented. Third, the main findings are reported. Finally, the results are discussed, and the limitations of this work and implications for future research are considered.

5.3 Theoretical background

Problem-solving

Problems arise when someone has a goal but lacks the knowledge of how to achieve it (Duncker, 1945). Various problem types exist. Jonassen (2000) provides a taxonomy of 11 problems from well-structured such as algorithmic problems, to ill-structured such as dilemmas, in which the more ill-structured problem types may encompass more structured problems. Following Dörner (1987; Funke, 2012), complex problems include various interconnected variables, multiple and conflicting goals, a lack of transparency, and dynamic development. It is difficult to accurately capture the scope of requirements to solve a particular problem. In educational contexts, problem difficulty is often assessed *ex post facto* based on solution rates (Jonassen & Hung, 2008). Furthermore, it is challenging to define the difficulty or complexity of a problem *a priori* because the subjective perception of a problem varies with prior knowledge and experience in the respective domain (Dörner, 1997). Similarly, Mayer and Wittrock (2006) distinguish routine and non-routine aspects of a problem, with a routine problem defined as a problem “for which the problem solver already possesses a ready-made solution procedure” (p. 288). In the context of competence assessment, Williamson et al. (2006) attempt to objectively define a task as complex if (a) the problem solver has to undergo multiple, non-trivial, domain-relevant steps and/or cognitive processes, (b) multiple features of task performance are captured, (c) task performance is relatively unconstrained, and (d) evaluations of task solutions recognise the interdependence of task features and aspects of performance.

Solving problems requires cognitive, metacognitive, and non-cognitive processes (Frensch & Funke, 1995; Jonassen, 2000; Mayer, 1998). Weinert (2001) defines such competencies as a combination of “[...] intellectual abilities, content-specific knowledge, cognitive skills, domain-

specific strategies, routines and subroutines, motivational tendencies, volitional control systems, personal value orientations, and social behaviours” (p. 51). Similarly, Fischer and Neubert (2015) define problem-solving competence as a multidimensional construct that includes knowledge, skills, abilities, and other components (KSAO), with ‘other components’ referring to non-cognitive facets such as frustration tolerance and a positive attitude in particular. Following Mayer and Wittrock (2006), problem-solving is preferably related to a specific domain instead of general heuristics (domain-specific principle), most likely restricted to a certain problem and not widely transferable to other problems (near-transfer principle), and should be integrated into teaching as guided problem-solving tasks to foster learning (knowledge integrating principle). Therefore, in order to promote problem-solving skills, problem-oriented tasks should be embedded as authentic, domain-specific scenarios in VET.

Problem-solving in computer-based simulations

Authentic domain-specific problems are so-called ‘metaproblems’ (Jonassen, 2000), a combination of many problem variations and types that are connected over a single domain. Metaproblems can be illustrated within computer-based simulations to replicate real-world tasks in a safe environment for training and learning purposes while providing an authentic and dynamic simulation-based learning scenario that changes either with decisions (interactions), with time, or both (Dörner & Funke, 2017). Such open-ended environments emphasize learner-centred activities, setting authentic tasks for learners, and providing them with authentic tools (Clarebout et al., 2009; Hannafin, 1995). The experimental learning opportunity addresses the cognitive, motivational, affective, psychomotor, and social aspects of learning (Breckwoldt et al., 2014). Early research on problem-solving was conducted within computer-simulated microworlds (Brehmer & Dörner, 1993). Nowadays, especially in the field of vocational education and training, there is a large number of domain-specific, authentic computer-based simulations to promote competence development in general and domain-specific problem-solving competence in particular (Beck et al., 2016; Rausch et al., 2016). However, most research in VET focuses on outcomes (competence assessment, learning performance, etc.) and not on the processes that precede these outcomes (Abele, 2018). Therefore, it seems worthwhile to highlight the methodological advantages of process data channels such as eye-tracking.

Eye-tracking, eye movements, and eye-tracking in problem-solving

Eye-tracking is a technology used as a research method for recording eye behaviour such as pupil dilation, blinking, and especially eye movements, as an indicator of visual attention when processing information (Duchowski, 2017; Holmqvist et al., 2011; Holmqvist & Andersson,

2017). Eye-tracking in computer-based simulations might help to make inferences about cognitive and metacognitive processes during learning (van Gog et al., 2009; van Gog & Jarodzka, 2013).

Eye movements reflect top-down (goal-driven or endogenous) and bottom-up (stimulus-driven or exogenous) visual attention (Orquin & Mueller Loose, 2013; Rayner, 1998; Theeuwes, 2010). Bottom-up control depends on stimulus features, such as visual saliency, i.e., the subjective quality of a stimulus that grabs visual attention (contrast, colour, movements). Top-down control depends on observer features, such as expertise, prior knowledge, tasks, etc. Also, individual eye features need to be considered. The most common types of eye movement events are fixations and saccades, where a fixation refers to the state when the eyes remain still (e.g., a stop during reading) while a saccade refers to the motion of the eyes between fixations (Holmqvist & Andersson, 2017). According to the ‘eye-mind hypothesis’, fixations should be a proxy of cognitive processing: “the eye remains fixated on a word as long as the word is being processed” (Just & Carpenter, 1980, p. 330). This influential assumption (originally related to reading research, but also tested beyond reading) has been challenged several times (Anderson et al., 2004; Underwood & Everatt, 1992), and today there is a consensus that visual attention somewhat precedes gaze and that overt and covert attention can differ (Holmqvist & Andersson, 2017).

Eye-movement data can reveal differences in visual attention to areas of interest (AOI) during problem-solving processes. Following Haider and Frensch’s ‘information reduction hypothesis’ (1996, 1999), deliberate practice helps students to learn to ignore redundant information and focus more on relevant information. Thus, especially experts have learned to distinguish relevant from irrelevant task-information through practice (F. J. Lee & Anderson, 2001) and make use of efficient cognitive strategies through experience (van Merriënboer, 2013). A meta-analysis examining the effects of expertise on visual comprehension conducted by Gegenfurtner and colleagues (2011) supported that experts have shorter fixation durations, more fixations on relevant areas, and fewer fixations on irrelevant areas than novices. Experts also showed selective attention through parafoveal processing (unattended locations of the visual field) indicated by longer saccades and shorter times to first fixation (Gegenfurtner et al., 2011). Additionally, research on eye-movement modelling examples (EMME) indicates that EMME might help to guide novices’ visual attention. EMME illustrates the visual processing behaviour of experts carefully performing a task by recordings their eye movements. A meta-analysis on EMME shows significant effects of eye-tracking measures such as time to first fixation and fixation

duration on novice learners' performance outcomes in terms of learning outcomes and problem-solving (Xie et al., 2021).

Despite the unquestioned potential of the eye-tracking approach, conclusions from eye-tracking data must be drawn very carefully. While there are a variety of eye movement measures, taxonomies, and interpretations, it is important to emphasize the theoretical assumptions as well as the domain and task-specific characteristics that underlie a research objective, and to our knowledge, there is currently no cross-domain taxonomy. However, a categorization based on operational definitions including a variety of eye-tracking measures can be used (see Holmqvist & Andersson, 2017; Part III on paradigms and measures). We would like to point out some aspects concerning the ambiguity of eye-tracking measurements. For example, if a participant shows a higher fixation duration for relevant information, this may indicate reasoning, but also confusion or mind-wandering ('staring into space'). Therefore, eye-tracking should be triangulated with offline and online data channels (which might induce other challenges to a research design), such as self-reports and self-assessments (which often suffer from several biases, Andrade, 2019), log files (which are often restricted to binary representations and suffers ambiguous interpretability, Goldhammer et al., 2014), which do not capture relevant off-screen behaviour (Maddox et al., 2018), as well as retrospective (delayed report) or concurrent (disturbing) think aloud (Gegenfurtner & Seppänen, 2013), and psychophysiological measures like heart rates which might be hard to interpret (Wu et al., 2014). However, the combination of multimodal data channels within advanced learning technologies is on the rise (Gabriel et al., 2022). Thus, the following review will also take additional behavioural measures into account to underline the potential of multimodal methods in combination with eye-tracking.

5.4 Methodology

Identification – Search strategy

A search and selection process in line with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020; see Liberati et al., 2009; Page et al., 2021) was conducted (Figure 5-2). The search was based on common educational research databases (Fraenkel et al., 2019): Web of Science, PsycInfo, and Education Resources Information Center (ERIC). We used the following four major search terms: (1) *eye-tracking AND (2) problem-solving AND (3) training and learning AND (4) computer-based simulation*. Eye-Tracking studies were identified by searching for *eye-tracking OR eye* OR gaze OR fixation OR saccade*. Problem-solving tasks were identified by searching for *problem-solving OR problem* OR decision-making OR decision* OR choice OR domain-specific* OR complex task*. The domains

of professional training or education were addressed by *education* OR vocation* OR apprentice* OR training OR program OR workshop OR workplace*. Finally, the search targets computer-based simulations and thus includes *computer OR simulation OR virtual*. The database search yielded 1,061 records. An additional check of reference lists for prominent eye-tracking reviews and meta-analysis was conducted (Gegenfurtner et al., 2011: a meta-analysis on expertise differences in visual comprehension; Lai et al., 2013: a review on ET studies in learning; Orquin & Mueller Loose, 2013: review of attentional shifts in decision tasks). Additionally, Google Scholar was manually searched with varieties of search strings. All findings were combined in a list, and duplicates were identified automatically. An additional manual check was performed. The search process resulted in 914 records for subsequent selection.

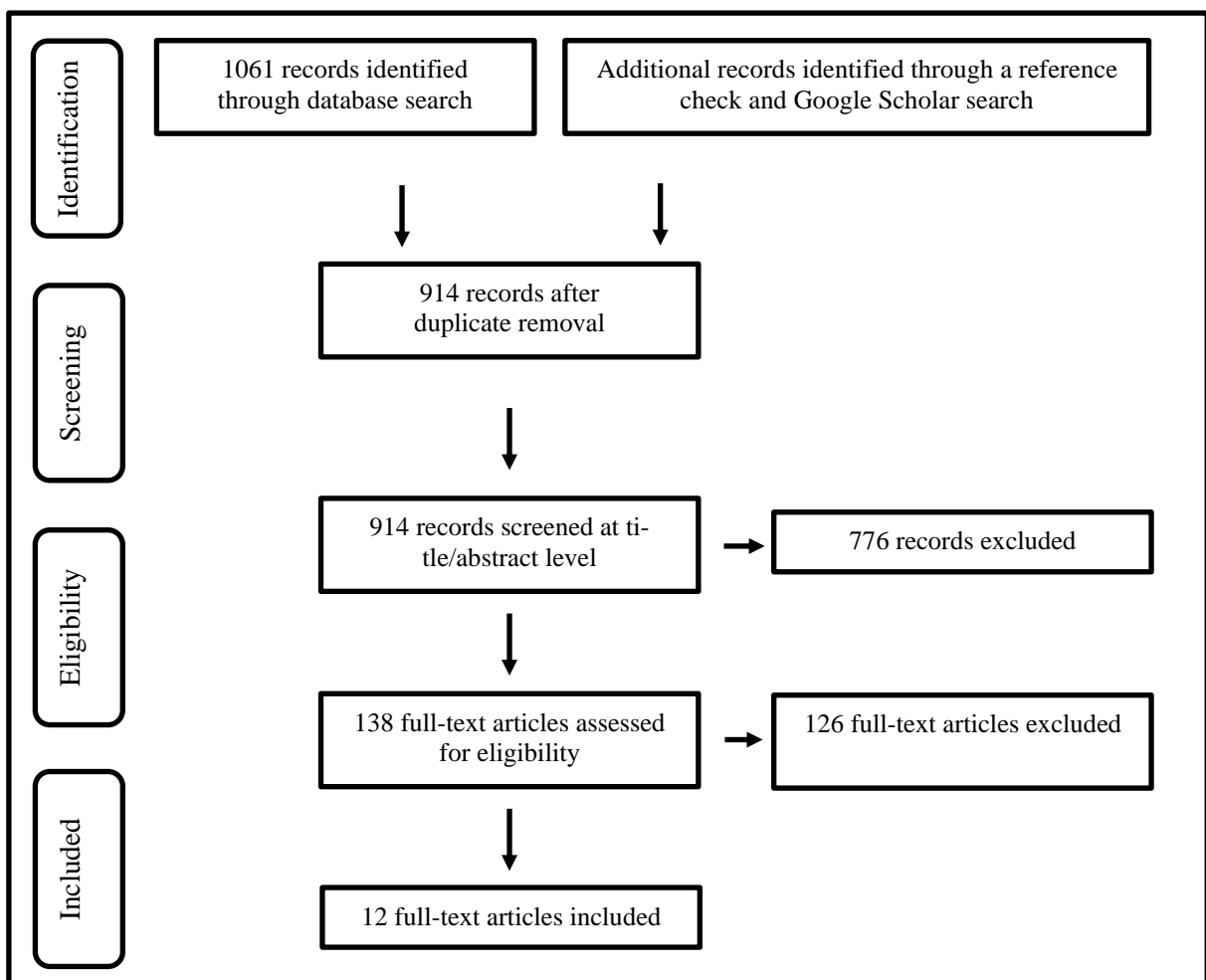


Figure 5-2: PRISMA flow chart.

Screening – Screening abstracts and titles

This review used CADIMA, a free web tool supporting the systematic review process (see Kohl et al., 2018). The selection criteria were applied at the title and abstract levels. Table 5-1 lists the key features, inclusion criteria, and exclusion examples.

Table 5-1: Key features, inclusion criteria, and exclusion examples.

Key features	Inclusion criteria	Exclusion examples
Participants	The study was conducted with healthy human beings as participants.	The study was conducted without humans (e.g., animals) or patients with mental or physiological disorders were subjects of interest.
Task	The study included a domain-specific problem-solving task.	The study included a cross-sectoral or non-authentic problem-solving task (e.g., Tower of Hanoi/London task).
Data collection	The study used eye-tracking as a primary method for data collection.	The study did not use eye-tracking as a method (e.g., mentioned eye-tracking for future research).
Type of publication	The article is an empirical study published in a peer-reviewed journal.	The article is a review paper, editorial note, or commentary.

We tested the inclusion criteria with a consistency check between two independent researchers to examine the inter-reviewer agreement and to overcome researcher bias. A second researcher screened randomly selected abstracts ($n = 45$; $>5\%$ of all potential records). The first check showed that the strength of agreement for screening (either inclusion, exclusion, or unsure) was $k = .60$ and considered as moderate (Landis & Koch, 1977). A kappa value of at least $k \geq .6$ is recommended to continue (Higgins et al., 2019; Orwin & Vevea, 2009). Further screening was revised by examining differences and resolving disagreements through discussion and debate by two reviewers, resulting in overall intercoder reliability of $k = .89$, which can be interpreted (Landis & Koch, 1977) as an almost perfect agreement ($k \geq .81$). Given the satisfying kappa value and for the sake of resource efficiency, the process proceeded with a single-person screening. A total of 914 records (after the elimination of duplicates), were screened. Overall, 776 records were excluded from this study.

Eligibility – Screening full texts

No papers were excluded for reasons of inaccessibility. If a paper was not available through institutional access, we kindly asked via ResearchGate® to provide the paper. A total of 138

records remained for screening at the full-text level. Overall, 126 full-text articles were excluded from the study. Owing to the high number of excluded articles, we provide some examples. We excluded studies that (1) did not conduct a complex problem-solving task or (2) were not constructed within an interactive environment.

(1): We excluded studies that did not match a domain-specific problem-oriented approach or interactive design. For example, we excluded a hypermedia-based learning environment that fostered self-regulated learning by using metacognitive pedagogical agents. The subjects learned about the human circulatory system for 90 minutes. However, a problem-solving task was not in the scope of this study (Taub & Azevedo, 2019). Studies based on a fixed stimulus without further interactions (mostly graphs or animations) were excluded. For example, we excluded a multiple-choice science study conducted by Tsai et al. (2011), which examines a science problem represented by four images in a web browser.

(2): We excluded studies on highly interactive learning environments for different reasons. A training simulator often integrates motor-sensitive skills. Training simulators contain realistic switches, knobs, levers, and typical instruments, e.g., flying simulations for pilot training (Schriver et al., 2008), and offshore drilling simulation (Naqvi et al., 2020). Participants train in dangerous scenarios to become familiar with safety processes in realistic but risk-free simulations and to develop routines. However, the comparability to computer-based learning simulations may not be given, because these training environments also involve motor skills and quick reactions in critical situations. Similarly, we excluded studies that included dummies or actors (O'Meara et al., 2015) and in situ experiments (Esau & Fletcher, 2018; Vrzakova et al., 2020). Finally, we excluded papers that were not related to the field of professional training or VET. On this basis, a total of only twelve articles could be included in the analysis.

Inclusion – Objectives of analysis and results

Included studies were analysed based on the domain, sample, task, performance measure, eye-tracking devices, measurements calculated, other behavioural measurements collected, analysis techniques, and main findings.

5.5 Results

Descriptives

All studies were published between 2005 and 2022, and seven of the studies were published in the last five years (Figure 5-3). Most studies were conducted in the United States (6), followed

by the Netherlands (3), Germany (1), Israel (1), and Taiwan (1). The sample size ranged between 7 and 70 participants ($M = 36.8$, $SD = 24.7$). The underlying domains are mainly related to the fields of science, technology, engineering, and mathematics (STEM), especially science (Cloude et al., 2020; Emerson et al., 2020; Taub et al., 2017), engineering (Abele et al., 2017; Gomes et al., 2013; van Gog, Paas, & van Merriënboer, 2005), but also the healthcare sector (Dubovi, 2022; J. Y. Lee et al., 2019, 2020) is represented (Table 5-2).

Table 5-2: Overview of domains.

Domain	Studies
Science	3
Technology	1
Engineering	3
Maths	0
Medical & Nursing	3
Air Traffic Control	1
Military	1

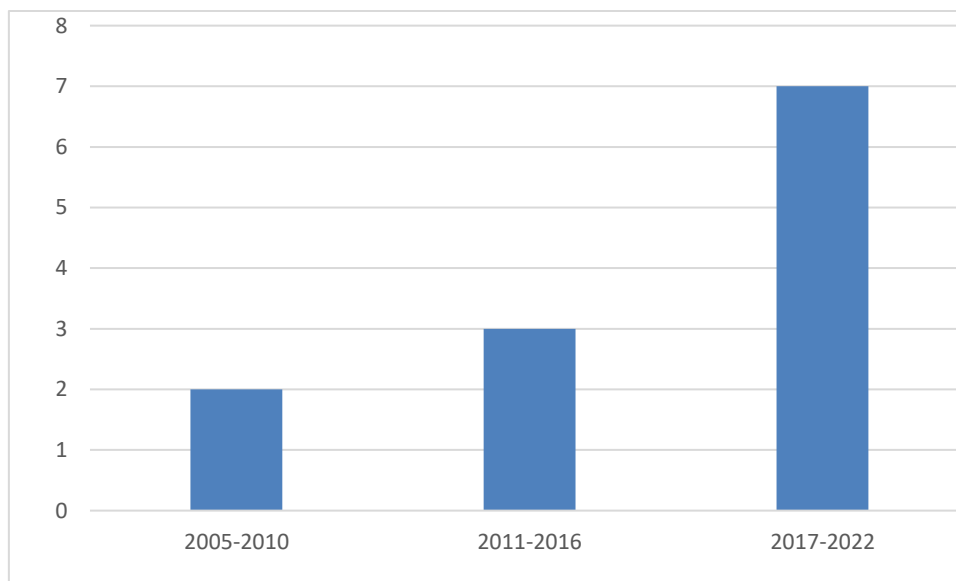


Figure 5-3: Included studies per 5-years interval.

Usually, students and/ or professionals were involved in the studies. Domain-specific complex problem-solving simulations include a broad variety of tasks and findings (see Appendix Table 5-4 and 5-5 for an overview). Typical tasks include problems like finding the cause of a mysterious disease outbreak (Cloude et al., 2020), fixing malfunctions in electrical circuits (van Gog, Paas, & van Merriënboer, 2005), or stabilizing a virtual patient by applying a medical

routine (J. Y. Lee et al., 2019). Performance was most often assessed manually (e.g., by counting the errors solved during troubleshooting), and also by examining pre-post-test scores (e.g., assessing content knowledge on microbiology), or through an automated log file analysis (e.g., by calculating a completion score for the underlying task).

Eye-tracking, measures, and analyses techniques

To answer RQ1, we analysed the different eye-tracking measures that were calculated and the most common analysis techniques that were applied. Further, we examined which additional behavioural measurements were collected next to eye-tracking data. Non-intrusive remote (also screen-based) eye-tracking devices were used in most of the studies (see Table 5-3 for an overview). Remote setups are typical for experimental lab studies, in contrast to eye-tracking glasses for field studies (e.g., Rosengrant et al., 2021) or virtual reality headsets (e.g. Torres et al., 2017). Currently, webcam-based eye-tracking is examined as a low-cost alternative next to remote eye-tracking devices (Wisiecka et al., 2022).

Table 5-3: Overview of eye-tracking setup, calculated measures, analysis techniques, and other collected measures.

Reference	Eye-tracking devices	Eye-tracking measures	Eye-tracking analyses	Add. offline/online measures
Abele et al. 2017	Remote (To-biiX120, 60Hz)	Total fixation duration (relevant AOIs)	Mann-Whitney-U	none
Cloude et al. 2020	Remote (SMI EYERED 250, 30Hz)	Proportion of total fixation time	Transformation to normal distribution, stepwise simple and multiple linear regression models by using AIC for selection	Log-file of mouse and keyboard data (event and time-based interaction)
Dubovi 2022	Remote (Smart Eye Aurora, 60Hz)	Time spent (relevant AOIs), time to first fixation, fixation counts, fixation dwells, saccade counts, blink rate	ANOVA, linear mixed effects model, prediction (two-phase approach): bivariate correlation with subsequent regression analysis	Facial expression recognition (Affdex), Self-report on affective state (PANAS), EDA (Shimmer 3),
Emerson et al. 2020	Remote (SMI RED 250, 120Hz)	Proportion of total fixation duration	ANOVA (feature selection), logistic regression (prediction)	Log file data (gameplay behaviour), facial expression recognition (FACET), self-reported Interest-Enjoyment (subscale of Intrinsic Motivation Inventory)

Reference	Eye-tracking devices	Eye-tracking measures	Eye-tracking analyses	Add. offline/online measures
Gomes et al. 2013	N/A	Number and duration per level on gaze points or fixations, direction of saccadic movement, type of pattern per trigram	k-means clustering	Log-file data (mouse clicks, time per level)
Kang and Landry 2014	Remote (Tobii X60)	Scan path	Scan path analysis according to simple orthogonal shapes	None
Lee et al. 2019	Remote (SMI RED, 250Hz)	Proportions of dwell time, fixation count, fixation duration, average fixation duration, fixation frequency, transition rate	MANOVA, t-tests, Mann-Whitney U	Log file data (Intervention completion score), mental workload (NASA-TLX), systematicity (Hidden Markov model; HMM)
Lee et al. 2020	Remote (SMI RED, 250Hz)	Pupil diameter, raw gaze points	Linear mixed effects model	Log file data (timestamps for in-game events such as pausing simulation), perceived mental effort.
Sohn et al. 2005	Head-mounted (ISCAN ETL-500)	Fixation time	F-test, power law fitting	None
Taub et al. 2017	Remote (SMI EYERED 250, 30 Hz)	Proportion of fixation duration	Multi-level modelling	Log file data (number of books read and reads per book, assessment submission attempts)
Tsai et al. 2016	Remote (faceLAB 4.6, 60Hz)	Percentage of time spent viewing (PTS), fixation duration (PFD), fixation count (PCD)	Mann-Whitney U, correlations, heatmaps, lag sequential analysis, heat map analysis per group	Self-reported flow experience

Reference	Eye-tracking devices	Eye-tracking measures	Eye-tracking analyses	Add. offline/online measures
(van Gog et al. 2005a)	Remote (SMI, 50Hz)	Fixation duration, number of fixations	Mann-Whitney U, Friedman test and Nemenyi post-hoc analysis, qualitative matching of verbal and gaze data.	Think-aloud (concurrent)

Following a functionally operational taxonomy for eye-tracking measures (Holmqvist et al., 2011; Holmqvist & Andersson, 2017), four types of measures are distinguished in this review: (1) movement measures: including movement measures of direction, amplitudes, duration, velocity, acceleration, shape, sequences and transitions, and scan path comparison measures; *ibid.*, p.439 ff., (2) position measures: including basic positions, measures of position dispersion, similarity, duration, and dilation; *ibid.*, p. 499 ff., (3) count measures: including a variety of countable entities such as saccades, smooth pursuits, blinks, fixations, dwells, AOIs, transitions, and more; *ibid.*, p. 574 ff., as well as (4) latency and distance measures: including the latency of a saccade, pupil dilation latency, eye-mouse distances, and more; *ibid.*, p. 579 ff.

Movement measures refer to the direction of eye movements and scan paths and are infrequently calculated. In total, three studies examined three different movement measures. Gomes et al. (2013) analysed the direction of saccadic movements (next to position measures) to examine eye movement patterns among high and low performers by applying a common unsupervised machine learning clustering (k-means) using trigrams of eye movements. Kang and Landry (2014) conducted a qualitative scan path analysis to examine whether novices in air traffic control follow a professional scan behaviour after being treated with expert scan path examples. Lee et al. (2019) calculated (next to several positions and count measures) the transition rates between AOIs to examine differences among experts and novices in their performance by applying t-test/Mann-Whitney-U and MANOVA.

Position measures are the most common measures used and refer to the positions where participants look. In total, position measures are calculated 15 times (in 11 of the 12 studies). Abele et al. (2017) measured the total fixation duration on relevant AOIs to analyze differences among performance groups by conducting a nonparametric Mann-Whitney-U test. Similar, Sohn et al. (2005) calculated fixation times to determine group differences. Cloude et al. (2020) calculated proportions of total fixation times to predict performance differences by applying stepwise simple and multiple linear regression models. Similarly, Emerson et al. (2020) integrated positional gaze data next to students' behavioural traces (such as gameplay behaviour and facial action

units) to predict performance and interest groups by running several logistic regression models with different feature compositions. Lee et al. (2020) examined the effects of pausing on the cognitive load within a medical serious game simulation by extracting pupil diameter and applying linear mixed effect models. A multi-level modelling approach was also applied by Taub et al. (2017) including proportions of fixation duration (next to interaction behaviour) to examine differences among performance while accounting for between and within-subject variances (random effects).

Count measures refer to the number and proportions of countable gaze behaviour and are frequently present. Count measures were calculated six times (3 studies) and were most often examined next to similar position measures. Tsai et al. (2016) calculated percentages of fixation counts (next to position measures) to analyse flow experience and visual attention among high and low-performing groups by applying the Mann-Whitney-U test. Also, AOI sequences were used for lag sequential analysis to examine different patterns of visual attention. Van Gog et al. (2005) examined the number of fixations (next to position measures) to examine expertise-related differences by applying Mann-Whitney U, the Friedman test with Nemenyi post-hoc analysis as well as a qualitative data analysis by matching verbal and gaze data.

Finally, latency and distance measures refer to time delays and space distance across eye movements and other points (e.g., mouse cursor). Dubovi (2022) examine latency and distance measures (next to position and count measures) by calculating time to the first fixation and applying ANOVA and linear mixed effect models for group and individual differences as well as regression analysis for performance predictions.

Eye-tracking data can be very ambiguous and is dependent on individual characteristics. To overcome this challenge, researchers are increasingly examining other behavioural data measurements in conjunction with eye-tracking data (Dewan et al., 2019) which can be used for data triangulation. Offline and online measures are frequently used next to eye-tracking (see Table 5-3). Log files (Cloude et al., 2020; Emerson et al., 2020; Gomes et al., 2013; J. Y. Lee et al., 2019, 2020) are often collected within computer-based simulations and result in additional and complementary insights into participants' behaviour through mouse clicks and keyboard strokes. Similarly, concurrent think-aloud can help to interpret eye-tracking data through the constant (or retrospective) verbalisation of participants' thoughts (van Gog, Paas, van Merriënboer, et al., 2005). Facial expression recognition (FER) algorithms analyse the expression (Dubovi, 2022; Emerson et al., 2020) of anger, disgust, fear, happiness, sadness, surprise, and underlying facial action units (mostly based on the Facial Action Coding System (FACS); Ekman & Friesen, 1976). Electrodermal activity (EDA) measures skin conduction as a proxy of

psychological or physiological arousal (Dubovi, 2022). Additionally, self-report questionnaires are used to measure subjective perceptions (Dubovi, 2022; Emerson et al., 2020; J. Y. Lee et al., 2019, 2020; Tsai et al., 2016).

Main findings

To answer RQ2, we group the main findings related to complex problem-solving in computer-based simulations based on process data measurements.

Most studies analyse differences across performance groups such as high and low performers or expert-novice comparisons and related patterns. Results show strong support for the information-reduction hypothesis following Haider and Frensch (1996, 1999). According to the information reduction hypothesis of Haider and Frensch (1996, 1999), deliberate practice helps learners to ignore redundant information and to focus more on relevant information. Thus, especially experts have learned through practice to distinguish relevant from irrelevant task information (F. J. Lee & Anderson, 2001) and to use efficient cognitive strategies due to prior experiences (van Merriënboer, 2013).

High performers or experts show a longer total fixation time and fewer fixations (Abele et al., 2017), higher proportions of dwell time to total time (with a large effect), a higher ratio of fixation count to total fixation counts (medium effect), and longer fixation duration (large effect) on critically relevant information (J. Y. Lee et al., 2019). High performers spend more time in a ‘problem orientation’ and ‘action evaluate & next action decision’ phase, they spend more fixations on fault-related components, show shorter mean fixation durations in an ‘orientation’ as well as longer mean fixation durations during a ‘formulation’ phase (van Gog, Paas, & van Merriënboer, 2005). It is also reported that through practice, less time is spent on relevant and irrelevant areas (Sohn et al., 2005). Similarly, novices performed better (made fewer false alerts), perceived an expert scan path as useful for their training, and tended to follow a professional expert scan pattern after treatment with an expert scan path (showed a circular movement across the air traffic control screen) (Kang & Landry, 2014). A combination of a shorter time to first fixation, fewer clicks, more unique fixations, and longer durations per fixation was found for the high-performance cluster (Gomes et al., 2013), while shorter durations for first fixations might indicate higher attentional readiness and indicates more time spent on reasoning before action. A longer time for the first fixation, a higher number of clicks and short fixation durations might indicate a lack of focus on the strategy or a lack of reasoning (trial and error) before action (Gomes et al., 2013).

In line with the results for experts and high performers, low performers or novices show shorter fixation times, more attention to similar but irrelevant AOIs, and lower visual accuracy. Low performers spend a higher proportion of time gathering information and less time generating hypotheses (Cloude et al., 2020). Low performers or novices show shorter or substantially longer fixation time (a behaviour that might indicate confusion) (Abele et al., 2017). The visual attention was spent on similar medicines, indicating processing difficulties through more fixation counts and dwells (Dubovi, 2022). The low comprehension group showed higher mental effort (Tsai et al., 2016) and paid more attention to graphic information (while a high comprehension group spent less on the graphical and more attention on the textual information) examined by qualitative heatmap analysis (Tsai et al., 2016).

Individual differences are related to prior knowledge and differences in cognitive load demands. Lower prior knowledge positively moderated the relation between interaction and fixation on gathering information in a serious game, while a negative relation for higher prior knowledge was found (Cloude et al., 2020). Less successful participants tended to get stuck in messages (cues) and an out-of-screen gaze while successful participants tended to transfer the knowledge and might use an out-of-screen gaze for pausing or reasoning (Tsai et al., 2016) to reduce cognitive load demands. Allowing pauses in a medical simulation increases the performance and cognitive load, regardless of whether pauses were taken or not. During pauses, the cognitive load was lower than during the simulation. When pauses were available, taking those pauses did not further benefit cognitive load or performance (J. Y. Lee et al., 2020). Pupillometry might be a valid measure of the cognitive load next to self-reports (J. Y. Lee et al., 2020).

Other online and offline process measures for behavioural data shed further light on differences in gaze behaviour when solving problems in computer-based simulations. Self-reports showed that a higher flow time distortion was associated with more fixations on the main task while lower flow time distortion was associated with fixations on the message prompts (Tsai et al., 2016). No significant changes in self-reported affective states over time were reported, while a higher level of presence was related to more visual attention to the relevant medicine (Dubovi, 2022). Facial expression recognition (FER) shows no significant impact of joy expression on post-tests, but frequent anger expressions were associated with lower post-test scores and positive emotions were related to inducing blinks (Dubovi, 2022). EDA shows a significant correlation between EDA peaks and blinks, but not with participants' emotional engagement (Dubovi, 2022). Eye-tracking data helps to supplement and contextualize log files (Cloude et al., 2020). Experts show higher levels of systematicity (indicated by the HMM score obtained through a log file analysis) (J. Y. Lee et al., 2019), and a negative effect between the number

of books and performance, as well as for the frequency of books and performance were found. The best performance was associated with reading fewer books but higher frequencies per book, emphasizing a quality reading strategy (fewer books more often) compared to a quantity reading (more books) strategy. Also, no unique association between proportions of fixations on book content or book concept matrix with individual submission attempts were found, but a significant interaction effect between both emphasizes the collection and combination of multichannel data. Low proportions of fixations on book content and concept matrices were related to high performance (Taub et al., 2017). Finally, concurrent think-aloud verbal data for high expertise participants show a predictive behaviour while low expertise participants' verbal data show no orientation and an unstructured initial testing approach.

Behavioural data might be further used for performance prediction within multimodal learning analytics. Gaze as a feature (unimodal) or gameplay, and face as a multimodal feature approach yields an accuracy of .67 for prediction among three performance groups, but adding more modalities comes at the cost of noise, so feature selection must be done carefully to avoid overfitting (Emerson et al., 2020). Also, gameplay and face (multimodal) yield .59 accuracy for prediction among three interest level groups (Emerson et al., 2020). The emotional and cognitive engagement measured via multimodal metrics explained 51% of post-test learning achievements (Dubovi, 2022). Interestingly, the blink rate is negatively associated with post-test scores and shows significantly lower rates during the actual problem (Dubovi, 2022). Significant associations between performance and the multimodal predictors as well as for the interaction term were found (Taub et al., 2017). The highest performance was related to a higher frequency of books, fewer books, and lower proportions of fixations on book content or concept matrix (Taub et al., 2017). Overall, multimodal data channels are very promising for further progress toward individualized learning analytics approaches (Cloude et al., 2020).

5.6 Discussion

This scoping review aimed to analyse the current state of eye-tracking research on domain-specific complex problem-solving in authentic tasks within interactive computer-based simulations. A total of twelve studies from a wide range of vocational education and professional training domains were found.

The most commonly calculated measures are position measures, and these are almost exclusively position duration measures such as the proportion of fixation time or total dwell time. Count measures are also mostly related to the number or proportion of fixations and dwell times. Surprisingly, movement measures are rarely computed and usually refer to saccade directions

or scan path. Heatmaps or scan paths are often qualitatively compared. There is a lack of quantitative approaches for measuring time patterns or similarity measurements for a scan path as stated by Holmqvist et al. (2011; 2017). Also, latency and distance measures are almost never calculated. This indicates that the potential to shed further light on complex problem-solving in computer-based simulation might not yet be fully exhausted by calculating other than the standard count and position duration eye-tracking measurements. The much broader variety of potential eye-tracking measures (concerning the underlying specific research questions) should be taken into account (Holmqvist et al., 2011; Holmqvist & Andersson, 2017). For example, cognitive load might be measured by considering saccadic peak velocity (Di Stasi et al., 2011), time to first fixations might be an indicator for visual attention to cues and hints in serious games (Conati et al., 2013), and saccade paths (Wu et al., 2014) might be calculated for further insights into behavioural differences and performance predictions.

To analyse eye-tracking data, group comparisons between experts and novices or high-performing and low-performing groups are often computed using common statistical methods such as t-test, (M)ANOVA, or non-parametric Mann-Whitney-U. Patterns between groups are examined with heat maps and lag sequential analyses, by mostly using discrete behaviour codes, or common k-means for clustering purposes. Recently, an increasing number of researchers have focused on individual differences in addition to the group level to account for random effects by applying linear mixed-effects models. This is relevant for eye-tracking research since eye movement data can vary between and within participants over time. The emphasis on the application of mixed-effects models in reading research to analyse eye-tracking data (Catrysse et al., 2018) shows further potential for generalizing results of between-group comparisons while accounting for within-subject variances, and additionally, increases the power of statistical analyses (compared to common approaches such as ANOVA) when conducted for lower aggregated levels (Baayen et al., 2008; Catrysse et al., 2018; Quené & van den Bergh, 2008). Finally, post-hoc performance predictions are the first attempts to develop multimodal learning analytics. However, these performance predictions are often performed as subsequent machine learning regression analyses, most of the reported accuracy scores seem to be currently not suited for practical implementation and are currently not integrated within the computer-based simulations for real-time assessments. Research out of the field of decision support systems (Causse et al., 2019) shows promising results to improve performance support.

Similarly, using multimodal data channels seems promising for educational purposes by integrating eye-tracking into systems of multimodal learning analytics (Cloude et al., 2022). In-

sights from additional questionnaires, think-aloud protocols, log files, and other psychophysiological measures have proven valuable next to eye-tracking data. Eye-tracking data combined with log file analysis and think-aloud protocols might be useful to validate each other and reveal further information about problem-solving processes (Stieff et al., 2011; van Gog, Paas, & van Merriënboer, 2005; van Gog, Paas, van Merriënboer, et al., 2005). Interestingly, Taub et al. (2017) found significant effects for the interaction term between gaze and log file data. They emphasize the use of multimodal data by stating “that our most significant results were those that included online trace data from both log files and eye tracking” (Taub et al., 2017, p. 651). In many cases, self-reporting is used as an additional measurement for data triangulation. In addition to eye-tracking, log files and facial expression recognition algorithms are also used. However, few studies use shimmer devices to detect electrodermal activity or practice concurrent thinking aloud. Studies on psychophysiological measures have shown to be valid indicators for problem-solving performance-related measures such as stress (Kärner et al., 2018). Self-reports and log files are useful tools for data triangulation. However, sometimes changes in affective state are not consciously perceived and reported through self-reports but can be measured through facial expression recognition algorithms, as reported by Dubovi (2022). Also, log files show higher systematicity for expert behaviour through HMM scores introduced by Lee et al. (2019). To obtain these systematicity scores, a rigorous task analysis must be performed before computation. Overall, despite the rise of multimodal approaches, the recognition of facial expressions using algorithms, measuring electrodermal activity using shimmer devices, and concurrent (or retrospective, for a comparison see van Gog, Paas, van Merriënboer, et al., 2005) thinking aloud are rare in this sample and data synchronisation remains a challenging aspect of research when data is not collected within a single software, which is not always possible (e.g., when log files of educational data are protected on a separate and secured server).

According to the “information reduction hypothesis” of Haider and Frensch (1996, 1999), deliberate practice helps learners to ignore redundant information and to focus more on relevant information. Thus, experts (and high performers) have learned through practice to distinguish relevant from irrelevant task information (F. J. Lee & Anderson, 2001) and to use efficient cognitive strategies through experience (van Merriënboer, 2013). This is also indicated by many studies in the sample. Similarly, low-performers or novices show shorter fixation times, more attention to similar but irrelevant AOIs, and lower visual accuracy. Performance in computer-based simulations and problem-solving seems to be moderated by prior knowledge, which positively influences the interaction between simulations and information fixation. Lower prior knowledge relates to lower performance and more fiddling around (trial and error strategy). The effects of pausing in simulations (for the medical field) were found to increase performance,

whether those pauses were taken or not. Also, successful problem solvers tended to take knowledge from cues when they were given, which could be related to the use of an out-of-screen gaze to pause or think, while unsuccessful participants got stuck in a loop between reading cues and an out-of-screen gaze.

Recently, some work has been done on the post hoc analysis of multimodal features such as eye-tracking and facial expression recognition data as well as log data for performance prediction. Interestingly, Emerson et al. (2020) state that that using more features for prediction comes at the cost of integrating more noise into the prediction, sometimes making a model's performance worse by overfitting. Some regression-based approaches seem promising and could explain up to 67% of the total variance. Nevertheless, it is difficult to determine machine learning performance by metrics such as accuracy. Especially for the more typical unbalanced datasets, other evaluation metrics such as F1 scores (the harmonic mean of precision and recall) are typically reported in machine learning research.

A major limitation of any literature review is publication bias. By addressing more than one database and a broad search term, as well as additional reference checks, we attempted to challenge publication bias appropriately. Despite these efforts, it is still possible that there is literature available but was not found. Though a review must aim to be all-inclusive, it may not always be possible. A single researcher performed most of the selection procedures. However, acceptable kappa values were calculated for a small number of studies between the two independent coders. Generalisability is not provided over a broad range of domains and tasks. There are shortcomings in the representation of countries and samples. Studies from Western countries are mainly represented within this sample. Also, we want to underline that high performers and experts are not the same (performance-based vs. criteria-based selection). A major shortcoming of this review is the limited number of studies analysed. Thus, we stick to a narrative scoping review but can give no information about the overall statistical effect due to sample restrictions and the heterogeneity in terms of study designs and dependent variables.

Future research might conduct a more systematic review and meta-analysis, particularly on the relationship between performance differences and eye movement measures. So far, within this specific subfield of interest, not enough studies were conducted and published to further examine such relationships. Thus, one advantage of this review is that we show the diversity of eye-tracking as a data collection method as well as different analysis techniques to foster eye-tracking research for VET domains, where computer-based simulations gain increasing relevance for education. This review supports eye-tracking as a data collection method for studying be-

havioural patterns in learning processes. Further studies should collect experiences during problem-solving processes (Rausch et al., 2019) and learning-related emotions by examining affective states through facial expression recognition (Munshi et al., 2020). Finally, there is a general research gap for eye-tracking studies and behavioural analysis in vocational education and training, and more precisely a vast lack of studies in the field of business education.

5.7 References

- Abele, S. (2018). Diagnostic Problem-Solving Process in Professional Contexts: Theory and Empirical Investigation in the Context of Car Mechatronics Using Computer-Generated Log-Files. *Vocations and Learning*, *11*(1), 133–159. <https://doi.org/10.1007/s12186-017-9183-x>
- Abele, S., Ostertag, R., Peissner, M., & Schuller, A. (2017). Eine Eye-Tracking-Studie zum diagnostischen Problemlöseprozess. Bedeutung der Informationsrepräsentation für den diagnostischen Problemlöseerfolg [An eye-tracking study on the diagnostic problem-solving process. The importance of information representation for diagnostic problem-solving success]. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik*, *113*(1), 86–109.
- Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, *125*, 413–428. <https://doi.org/10.1016/j.compedu.2018.06.023>
- Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. *Psychological Science*, *15*(4), 225–231.
- Andrade, H. L. (2019). A Critical Review of Research on Student Self-Assessment. *Frontiers in Education*, *4*, 87. <https://doi.org/10.3389/educ.2019.00087>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Ashraf, H., Sodergren, M. H., Merali, N., Mylonas, G., Singh, H., & Darzi, A. (2018). Eye-tracking technology in medical education: A systematic review. *Medical Teacher*, *40*(1), 62–69. <https://doi.org/10.1080/0142159X.2017.1391373>
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Beck, K., Landenberger, M., & Oser, F. (Eds.). (2016). *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der BMBF-Förderinitiative ASCOT [Technology-based skills assessment in vocational education and training: results from the BMBF funding initiative ASCOT]*. W. Bertelsmann.
- Breckwoldt, J., Gruber, H., & Wittmann, A. (2014). Simulation Learning. In S. Billett, C. Harteis, & H. Gruber (Eds.), *International Handbook of Research in Professional and Practice-based Learning* (pp. 673–698). Springer. https://doi.org/10.1007/978-94-017-8902-8_25
- Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, *9*(2–3), 171–184. [https://doi.org/10.1016/0747-5632\(93\)90005-D](https://doi.org/10.1016/0747-5632(93)90005-D)

- Catrysse, L., Gijbels, D., Donche, V., De Maeyer, S., Lesterhuis, M., & Van den Bossche, P. (2018). How are learning strategies reflected in the eyes? Combining results from self-reports and eye-tracking. *British Journal of Educational Psychology*, 88(1), 118–137. <https://doi.org/10.1111/bjep.12181>
- Causse, M., Lancelot, F., Maillant, J., Behren, J., Cousy, M., & Schneider, N. (2019). Encoding decisions and expertise in the operator's eyes: Using eye-tracking as input for system adaptation. *International Journal of Human-Computer Studies*, 125, 55–65. <https://doi.org/10.1016/j.ijhcs.2018.12.010>
- Clarebout, G., Elen, J., Lowyck, J., Van den Ende, J., & Van den Enden, E. (2009). Tropical Medicine Open Learning Environment: In P. L. Rogers, G. A. Berg, J. V. Boettcher, C. Howard, L. Justice, & K. D. Schenk (Eds.), *Encyclopedia of Distance Learning, Second Edition* (pp. 2155–2159). IGI Global. <https://doi.org/10.4018/978-1-60566-198-8.ch318>
- Cloude, E. B., Azevedo, R., Winne, P. H., Biswas, G., & Jang, E. E. (2022). System design for using multimodal trace data in modeling self-regulated learning. *Frontiers in Education*, 7, 928632. <https://doi.org/10.3389/educ.2022.928632>
- Cloude, E. B., Dever, D. A., Wiedbusch, M. D., & Azevedo, R. (2020). Quantifying Scientific Thinking Using Multichannel Data With Crystal Island: Implications for Individualized Game-Learning Analytics. *Frontiers in Education*, 5, 572546. <https://doi.org/10.3389/educ.2020.572546>
- Conati, C., Jaques, N., & Muir, M. (2013). Understanding Attention to Adaptive Hints in Educational Games: An Eye-Tracking Study. *International Journal of Artificial Intelligence in Education*, 23, 136–161. <https://doi.org/10.1007/s40593-013-0002-8>
- Dewan, M. A. A., Murshed, M., & Lin, F. (2019). Engagement detection in online learning: A review. *Smart Learning Environments*, 6, 1–20. <https://doi.org/10.1186/s40561-018-0080-z>
- Di Stasi, L., Antoli, A., & Canas, J. (2011). Main sequence: An index for detecting mental workload variation in complex tasks. *Applied Ergonomics*, 42(6), 807–813. <https://doi.org/10.1016/j.apergo.2011.01.003>
- Dörner, D. (1987). *Problemlösen als Informationsverarbeitung [Problem solving as information processing]*. Kohlhammer.
- Dörner, D. (1997). *The logic of failure: Recognizing and avoiding error in complex situations*. Basic Books.
- Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, 8, 1153. <https://doi.org/10.3389/fpsyg.2017.01153>
- Dubovi, I. (2022). Cognitive and emotional engagement while learning with VR: The perspective of multimodal methodology. *Computers & Education*, 183, 104495. <https://doi.org/10.1016/j.compedu.2022.104495>
- Duchowski, A. T. (2017). *Eye Tracking Methodology*. Springer. <https://doi.org/10.1007/978-3-319-57883-5>
- Duncker, K. (1945). On problem-solving. (L. S. Lees, Trans.). *Psychological Monographs*, 58(5), i–113. <https://doi.org/10.1037/h0093599>
- Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, 1(1), 56–75.

- Emerson, A., Cloude, E. B., Azevedo, R., & Lester, J. (2020). Multimodal learning analytics for game-based learning. *British Journal of Educational Technology*, *51*(5), 1505–1526. <https://doi.org/10.1111/bjet.12992>
- Esau, T., & Fletcher, S. (2018). Prozessorientierte Analyse von konstruktiven Problemlöseprozessen auf Basis von Eye-Tracking-Aufnahmen [Process-oriented analysis of constructive problem-solving processes on the basis of eye-tracking recordings]. *Journal of Technical Education*, *6*(1), 2198–0306. <https://doi.org/10.48513/joted.v6i1.116>
- Fischer, A., & Neubert, J. C. (2015). The multiple faces of complex problems: A model of problem solving competency and its implications for training and assessment. *Journal of Dynamic Decision Making*, *1*, 1–13. <https://doi.org/10.11588/jddm.2015.1.23945>
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education*. McGraw Hill.
- Frensch, P. A., & Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving. The European perspective* (pp. 3–25). Lawrence Erlbaum.
- Funke, J. (2012). Complex Problem Solving. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 682–685). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_685
- Gabriel, F., Cloude, E. B., & Azevedo, R. (2022). Using Learning Analytics to Measure Motivational and Affective Processes During Self-Regulated Learning with Advanced Learning Technologies. In Y. “Elle” Wang, S. Joksimović, M. O. Z. San Pedro, J. D. Way, & J. Whitmer (Eds.), *Social and Emotional Learning and Complex Skills Assessment* (pp. 93–108). Springer. https://doi.org/10.1007/978-3-031-06333-6_6
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains. *Educational Psychology Review*, *23*(4), 523–552. <https://doi.org/10.1007/s10648-011-9174-7>
- Gegenfurtner, A., & Seppänen, M. (2013). Transfer of expertise: An eye tracking and think aloud study using dynamic medical visualizations. *Computers & Education*, *63*, 393–403. <https://doi.org/10.1016/j.compedu.2012.12.021>
- Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, *106*(3), 608–626. <https://doi.org/10.1037/a0034716>
- Gomes, J. S., Yassine, M., Worsley, M., & Blikstein, P. (2013). Analysing Engineering Expertise of High School Students Using Eye Tracking and Multimodal Learning Analytics. In S. D’Mello, R. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 1–3).
- Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology*, *30*(3), 304–337.
- Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(1), 172–190. <https://doi.org/10.1037/0278-7393.25.1.172>
- Hannafin, M. J. (1995). Open-Ended Learning Environments: Foundations, Assumptions, and Implications for Automated Design. In R. D. Tennyson & A. E. Barron (Eds.), *Automating Instructional Design: Computer-Based Development and Delivery Tools* (pp. 101–129). Springer. https://doi.org/10.1007/978-3-642-57821-2_5

- Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. Wiley.
- Holmqvist, K., & Andersson, R. (2017). *Eye tracking: A comprehensive guide to methods, paradigms, and measures* (2nd edition). CreateSpace.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- Jarodzka, H. (2021). Research Methods in Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 41–54). Cambridge University Press. <https://doi.org/10.1017/9781108894333.006>
- Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in Educational Science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research*, 10(1), 1–18. <https://doi.org/10.16910/JEMR.10.1.3>
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63–85. <https://doi.org/10.1007/BF02300500>
- Jonassen, D. H., & Hung, W. (2008). All Problems are Not Equal: Implications for Problem-Based Learning. *Interdisciplinary Journal of Problem-Based Learning*, 2(2), 6–28. <https://doi.org/10.7771/1541-5015.1080>
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. <https://doi.org/10.1037/0033-295X.87.4.329>
- Kang, Z., & Landry, S. J. (2014). Using scanpaths as a learning method for a conflict detection task of multiple target tracking. *Human Factors*, 56(6), 1150–1162. <https://doi.org/10.1177/0018720814523066>
- Kärner, T., Minkley, N., Rausch, A., Schley, T., & Sembill, D. (2018). Stress and Resources in Vocational Problem Solving. *Vocations and Learning*, 11(3), 365–398. <https://doi.org/10.1007/s12186-017-9193-8>
- Kohl, C., McIntosh, E. J., Unger, S., Haddaway, N. R., Kecke, S., Schiemann, J., & Wilhelm, R. (2018). Online tools supporting the conduct and reporting of systematic reviews and systematic maps: A case study on CADIMA and review of existing tools. *Environmental Evidence*, 7, 8. <https://doi.org/10.1186/s13750-018-0115-5>
- Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, 10, 90–115. <https://doi.org/10.1016/j.edurev.2013.10.001>
- Lajoie, S. P., & Naismith, L. (2012). Computer-Based Learning Environments. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 716–718). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_512
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Lee, F. J., & Anderson, J. R. (2001). Does learning a complex task have to be complex?: A study in learning decomposition. *Cognitive Psychology*, 42(3), 267–316.
- Lee, J. Y., Donkers, J., Jarodzka, H., & Merrienboer, J. van. (2019). How prior knowledge affects problem-solving performance in a medical simulation game: Using game-logs and eye-tracking. *Computers in Human Behavior*, 99, 268–277. <https://doi.org/10.1016/j.chb.2019.05.035>

- Lee, J. Y., Donkers, J., Jarodzka, H., Sellenraad, G., & van Merriënboer, J. J. G. (2020). Different effects of pausing on cognitive load in a medical simulation game. *Computers in Human Behavior, 110*, 106385. <https://doi.org/10.1016/j.chb.2020.106385>
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ (Clinical Research Ed.), 339*, b2700. <https://doi.org/10.1136/bmj.b2700>
- Maddox, B., Bayliss, A. P., Fleming, P., Engelhardt, P. E., Edwards, S. G., & Borgonovi, F. (2018). Observing response processes with eye tracking in international large-scale assessments: Evidence from the OECD PIAAC assessment. *European Journal of Psychology of Education, 33*(3), 543–558. <https://doi.org/10.1007/s10212-018-0380-2>
- Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science, 26*(1/2), 49–63. <https://doi.org/10.1023/A:1003088013286>
- Mayer, R. E., & Wittrock, M. C. (2006). Problem Solving. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of Educational Psychology* (2nd ed., pp. 287–303). Routledge.
- Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology, 18*(1), 143. <https://doi.org/10.1186/s12874-018-0611-x>
- Munshi, A., Mishra, S., Zhang, N., Paquette, L., Ocumpaugh, J., Baker, R., & Biswas, G. (2020). Modeling the Relationships Between Basic and Achievement Emotions in Computer-Based Learning Environments. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education* (Vol. 12163, pp. 411–422). Springer International Publishing. https://doi.org/10.1007/978-3-030-52237-7_33
- Naqvi, S., Raza, M., Ghazal, S., Salehi, S., Kang, Z., & Teodoriu, C. (2020). Simulation-based training to enhance process safety in offshore energy operations: Process tracing through eye-tracking. *Process Safety and Environmental Protection, 138*, 220–235. <https://doi.org/10.1016/j.psep.2020.03.016>
- O'Meara, P., Munro, G., Williams, B., Cooper, S., Bogossian, F., Ross, L., Sparkes, L., Browning, M., & McClounan, M. (2015). Developing situation awareness amongst nursing and paramedicine students utilizing eye tracking technology and video debriefing techniques: A proof of concept paper. *International Emergency Nursing, 23*(2), 94–99. <https://doi.org/10.1016/j.ienj.2014.11.001>
- Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica, 144*(1), 190–206. <https://doi.org/10.1016/j.actpsy.2013.06.003>
- Orwin, R. G., & Vevea, J. L. (2009). Evaluating coding decisions. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 177–203). Russell Sage Foundation.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. <https://doi.org/10.1136/bmj.n71>

- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413–425. <https://doi.org/10.1016/j.jml.2008.02.002>
- Rausch, A., Kögler, K., & Seifried, J. (2019). Validation of Embedded Experience Sampling (EES) for Measuring Non-cognitive Facets of Problem-Solving Competence in Scenario-Based Assessments. *Frontiers in Psychology*, 10, 1200. <https://doi.org/10.3389/fpsyg.2019.01200>
- Rausch, A., Seifried, J., Wuttke, E., Kögler, K., & Brandt, S. (2016). Reliability and validity of a computer-based assessment of cognitive and non-cognitive facets of problem-solving competence in the business domain. *Empirical Research in Vocational Education and Training*, 8(1), 9. <https://doi.org/10.1186/s40461-016-0035-y>
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.
- Rosengrant, D., Herrington, D., & O'Brien, J. (2021). Investigating Student Sustained Attention in a Guided Inquiry Lecture Course Using an Eye Tracker. *Educational Psychology Review*, 33(1), 11–26. <https://doi.org/10.1007/s10648-020-09540-2>
- Schrivver, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise differences in attentional strategies related to pilot decision making. *Human Factors*, 50(6), 864–878. <https://doi.org/10.1518/001872008X374974>
- Sohn, M., Douglass, S., Chen, M., & Anderson, J. (2005). Characteristics of fluent skills in a complex, dynamic problem-solving task. *Human Factors*, 47(4), 742–752. <https://doi.org/10.1518/001872005775570943>
- Stieff, M., Hegarty, M., & Deslongchamps, G. (2011). Identifying Representational Competence with Multi-Representational Displays. *Cognition and Instruction*, 29(1), 123–145. <https://doi.org/10.1080/07370008.2010.507318>
- Strohmaier, A. R., Schiepe-Tiska, A., Chang, Y.-P., Müller, F., Lin, F.-L., & Reiss, K. M. (2020). Comparing Eye Movements during Mathematical Word Problem Solving in Chinese and German. *ZDM: The International Journal on Mathematics Education*, 52(1), 45–58. <https://doi.org/10.1007/s11858-019-01080-6>
- Taub, M., & Azevedo, R. (2019). How Does Prior Knowledge Influence Eye Fixations and Sequences of Cognitive and Metacognitive SRL Processes during Learning with an Intelligent Tutoring System? *International Journal of Artificial Intelligence in Education*, 29(1), 1–28. <https://doi.org/10.1007/s40593-018-0165-4>
- Taub, M., Mudrick, N. V., Azevedo, R., Millar, G. C., Rowe, J., & Lester, J. (2017). Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with Crystal Island. *Computers in Human Behavior*, 76, 641–655. <https://doi.org/10.1016/j.chb.2017.01.038>
- Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, 135(2), 77–99. <https://doi.org/10.1016/j.actpsy.2010.02.006>
- Torres, F., Neira Tovar, L. A., & del Rio, M. S. (2017). A Learning Evaluation for an Immersive Virtual Laboratory for Technical Training Applied into a Welding Workshop. *EURASIA Journal of Mathematics, Science & Technology Education*, 13(2), 521–532.
- Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., & Yang, F.-Y. (2011). Visual Attention for Solving Multiple-Choice Science Problem: An Eye-Tracking Analysis. *Computers & Education*, 58(1), 375–385. <https://doi.org/10.1016/j.compedu.2011.07.012>

- Tsai, M.-J., Huang, L.-J., Hou, H.-T., Hsu, C.-Y., & Chiou, G.-L. (2016). Visual behavior, flow and achievement in game-based learning. *Computers & Education*, 98, 115–129. <https://doi.org/10.1016/j.compedu.2016.03.011>
- Underwood, G., & Everatt, J. (1992). The Role of Eye Movements in Reading: Some Limitations of the Eye-Mind Assumption. In E. Chekaluk & K. Llewellyn (Eds.), *The Role of Eye Movements in Perceptual Processes* (pp. 111–169). Elsevier. [https://doi.org/10.1016/S0166-4115\(08\)61744-6](https://doi.org/10.1016/S0166-4115(08)61744-6)
- van Gog, T., & Jarodzka, H. (2013). Eye Tracking as a Tool to Study and Enhance Cognitive and Metacognitive Processes in Computer-Based Learning Environments. In R. Azevedo & V. Aleven (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 143–156). Springer. https://doi.org/10.1007/978-1-4419-5546-3_10
- van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25(3), 785–791. <https://doi.org/10.1016/j.chb.2009.02.007>
- van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data: Uncovering expertise-related differences. *Applied Cognitive Psychology*, 19(2), 205–221. <https://doi.org/10.1002/acp.1112>
- van Gog, T., Paas, F., van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting. *Journal of Experimental Psychology: Applied*, 11(4), 237–244. <https://doi.org/10.1037/1076-898X.11.4.237>
- van Merriënboer, J. J. G. (2013). Perspectives on problem solving and instruction. *Computers & Education*, 64, 153–160. <https://doi.org/10.1016/j.compedu.2012.11.025>
- Van Ostaeyen, S., Embo, M., Schellens, T., & Valcke, M. (2022). Training to Support ePortfolio Users During Clinical Placements: A Scoping Review. *Medical Science Educator*, 32(4), 921–928. <https://doi.org/10.1007/s40670-022-01583-0>
- Vrzakova, H., Begel, A., Mehtätalo, L., & Bednarik, R. (2020). Affect Recognition in Code Review: An In-situ Biometric Study of Reviewer's Affect. *Journal of Systems and Software*, 159, 110434.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Hogrefe.
- Williamson, D. M., Mislevy, R. J., & Bejar, I. I. (2006). Automated scoring of complex tasks in computer-based testing: An Introduction. In D. M. Williamson, R. J. Mislevy, & I. I. Bejar (Eds.), *Automated scoring of complex tasks in computer-based testing* (pp. 1–13). Psychology Press.
- Wisiecka, K., Krejtz, K., Krejtz, I., Sromek, D., Cellary, A., Lewandowska, B., & Duchowski, A. (2022). Comparison of Webcam and Remote Eye Tracking. *2022 Symposium on Eye Tracking Research and Applications*, 1–7. <https://doi.org/10.1145/3517031.3529615>
- Wu, C.-H., Tzeng, Y.-L., & Huang, Y. M. (2014). Understanding the relationship between physiological signals and digital game-based learning outcome. *Journal of Computers in Education*, 1(1), 81–97. <https://doi.org/10.1007/s40692-014-0006-x>
- Xie, H., Zhao, T., Deng, S., Peng, J., Wang, F., & Zhou, Z. (2021). Using eye movement modelling examples to guide visual attention and foster cognitive performance: A meta-analysis. *Journal of Computer Assisted Learning*, 37(4), 1194–1206. <https://doi.org/10.1111/jcal.12568>

- Yang, F.-Y., Tsai, M.-J., Chiou, G.-L., Lee, S. W.-Y., Chang, C.-C., & Chen, L.-L. (2018). Instructional suggestions supporting science learning in digital environments based on a review of eye tracking studies. *Journal of Educational Technology & Society*, 21(2), 28–45.

6 Visual attention while solving economic learning tasks using dynamic, interactive graphs – An eye-tracking study

Study 4 is prepared for submission in ‘The Journal of Economic Education’.¹²

6.1 Abstract

In economic education, price-quantity graphs are often used to visualize the interplay of supply and demand. Research in multimedia learning in general and learning with visualizations in particular is widespread. However, research in the domain of economic education focusing on learning with static versus dynamic, interactive graphs is sparse. Interacting with dynamic graphs might increase engagement with the content and therefore foster learning. We investigate how economic education students ($n = 31$) at a German university learn with static and dynamic graphs by analyzing their gaze behavior. We use eye tracking while students are solving single-choice learning tasks with price-quantity graphs. Analyzing students’ visual attention and task performance, we found that their mean performance was higher on tasks where they could use dynamic graphs compared to tasks where they worked with static graphs. Differences between static and dynamic graphs in students’ visual attention were found with respect to average fixation duration. In addition, total fixation duration predicted performance on exercises with dynamic graphs (when controlling for students’ prior knowledge and attitude towards economic models). Our results suggest that dynamic graphs support knowledge retrieval better than static graphs in a learning environment for economic education.

Keywords: technology-enhanced learning; learning with visualizations; economic education; eye tracking

6.2 Introduction

Learning designs for economic education have changed, partly in response to the COVID-19 pandemic, and have become increasingly digital (Chaudhury, 2023). Price-quantity graphs are widely used in the classroom to teach basic economic principles at all levels of upper-secondary education (Friebel-Piechotta, 2022), for instance, to illustrate the relationships between supply and demand, to demonstrate how the price mechanism works under certain conditions (e.g., for different market structures, under different elasticity assumptions or in the presence of external

¹² Note: The chapter headings, figures, and tables of the published version have been numbered according to the numbering of this thesis. Minor changes or deviations from the final publication are possible. Spelling and grammar have been adjusted to American English standards. Any misspellings and typographical errors detected in the course of a further thorough proofreading have been corrected.

shocks), or to explain the welfare implications of economic decisions (e.g., to discuss changes in consumer and producer surplus due to government policies). Learning materials, including price-quantity graphs, are designed to visualize relationships in the neoclassical supply and demand model and to help learners build a mental representation of basic economic concepts. Supply and demand functions are both represented as a line (or curve) on a static graph that shows the aggregated behavior of buyers and sellers in a market for goods and services with prices and quantities on the axis. Changes in the market can cause shifts in supply or demand, and the graph helps visualize the effects of different conditions on market equilibria, prices, and quantities.

In digital learning environments, graphs can be implemented in ways that allow for an interactive engagement with dynamically displayed content. These dynamic contents are particularly beneficial for learning when learners can interact with and control the presentation (Tversky et al., 2002). Compared to static formats, dynamic ones, such as videos and animations, are particularly good at illustrating change over time (Ploetzner & Lowe, 2004). There is an ongoing debate about whether static graphics or animations are more suitable for learning and which conditions, e.g., design or learner characteristics, could be relevant (Zhuang & Liu, 2022). The learning environment used in this study, enhanced by both static and dynamic, interactive graphs within the field of economic education, was designed to improve economic knowledge by allowing students to interact with dynamic graphs (Findeisen et al., 2022). Working with dynamic interactive graphs can increase engagement with the material and therefore support learning processes (e.g., Rodgers et al., 2023). However, dynamic interactive graphs also place greater demands on the working memory, e.g. when redundancies occur (Kalyuga & Sweller, 2014). Although the use of dynamic graphs in instructional design is widespread, mixed results have been found in the past. Dynamic graphs can be useful in learning but may be more effective when it comes to motor tasks or mimicking human movements (Höffler & Leutner, 2007; Van Gog et al., 2009). The visual processing of graphs during problem solving and learning is an ongoing topic in research (Ruf et al., 2023).

Studies show that young adults lack economic literacy not only in the U.S. (Walstad & Rebeck, 2001) but also in Germany (Happ, Förster, et al., 2016; Happ, Zlatkin-Troitschanskaia, et al., 2016). Hence, understanding students' learning processes and identifying instructional elements that foster learning in economics is of high importance. The aim of this study is to investigate differences in visual attention when solving economic learning tasks in a learning environment with static and dynamic, interactive price-quantity graphs. In this regard, we contribute

to a better understanding of the added value of interactive dynamic graphics compared to static graphs.

6.3 Theoretical background

In multimedia learning, i.e., learning from words and pictures (Mayer, 2021), visualizations are a critical component alongside verbal or written instruction to support and enhance learning in a variety of subjects and professions. Task-supporting visualizations can illustrate concepts and relationships to learners in a different way than words and formulas, helping learners build multiple mental representations (Schnotz & Bannert, 2003). According to the cognitive theory of multimedia learning, learning environments designed in the way the human mind works can lead to more successful learning, i.e., the acquisition of integrated and transferable knowledge, as opposed to rote learning (Mayer, 2021). The theory draws on several cognitive theories and is based on three assumptions. First, the cognitive theory of multimedia learning is based on the dual coding theory (Paivio, 1990), which states that humans process visual and auditory information in separate channels (Sorden, 2013). Second, according to the cognitive load theory (Sweller et al., 1998, 2019) and the working memory model (Baddeley, 1997), the amount of information that can be processed in each channel is thereby limited in its capacity (Sorden, 2013). Finally, during learning, the human mind is assumed to be an active learner, engaged in processing information by selecting, organizing, and integrating presented material to construct mental representations (Mayer, 2021; Sorden, 2013).

Individual learners' characteristics, such as psychological, behavioral, and attitudinal aspects, affect the learning process (Nakayama & Santiago, 2012). Attitude is characterized as a predisposition or tendency to evaluate something in a particular way (Seel, 2012), representing the degree to which an individual has a favorable or unfavorable view of something (e.g., attitude towards ICT or economic models). Moreover, a learner's working memory has a significant impact on the learning process. However, the most important factor influencing learning is prior knowledge (Kalyuga, 2013). Multimedia presentations that are effective for learners with little knowledge may be less effective as learners acquire more specialized knowledge.

Among other online measures of the learning process (e.g. concurrent think-aloud, observations, brain imaging techniques), eye-tracking is an unobtrusive and suitable approach for recording individual learning behavior. When analyzing visual attention in multimedia learning environments, eye tracking can provide valuable insights for the design of computer-based learning and assessment environments (Jarodzka et al., 2017). Eye tracking is a research method to record eye movements (Duchowski, 2017; Holmqvist et al., 2011). The most common basic

eye movements are fixations and saccades. A fixation refers to the state when the eyes are still, while a saccade refers to the movement of the eyes between fixations. The cognitive processes underlying multimedia learning can be associated with eye movement measures (Alemdag & Cagiltay, 2018). For example, selecting information can be associated with the time of the first fixation on a particular area of interest (AOI), while organizing information can be associated with the average fixation duration on an AOI, and integrating can be associated with the number of transitions between text and images. Visual attention to organize information is considered a critical part of learning. This is because longer fixations are generally associated with deeper processing (Holmqvist & Andersson, 2017) and therefore indicate greater engagement with the material.

6.4 State of research

Learning effectiveness of visualizations

Since the 1990s, static diagrams have been widely used to teach economic principles, and the effects of presented and self-drawn diagrams on learning have been studied in economics education (Cohn & Cohn, 1994; Hoyt & McGoldrick, 2019), along with algebraic formulas (Zetland et al., 2010) and verbal instructions (Kourilsky & Wittrock, 1987). Empirical findings on the promotion of learning through graphs presented by the teacher are unclear (Cohn et al., 2001). The extent to which basic economic principles can be acquired when learning with graphs seems to depend on individual differences and can be facilitated by prior knowledge, mathematical preferences and attitudes towards graphs (Cohn et al., 2004; Hill & Stegner, 2003). In addition, students often find it difficult to draw accurate graphs (Cohn & Cohn, 1994), so they may not benefit from drawing graphs. While advanced teaching methods, like simulations of the dynamic behavior of economic systems (Sierra, 2020; Wheat, 2009), help to demonstrate complexity, it is essential to first understand the fundamental principles.

The learning effectiveness of static images versus animations has been studied in educational research for some time (Rieber, 1991). Animations are particularly conducive to learning when learners can interactively engage and control the representation themselves (Tversky et al., 2002). A meta-analysis by Höffler and Leutner (2007) showed an overall substantial advantage ($d = 0.37$) of animations over static pictures. Similarly, Berney and Bétrancourt (2016) found an overall positive effect of animations over static graphics ($g = 0.23$). In addition, the positive effect was moderated when the speed was predetermined by the system, when it was coupled with auditory comments, and when the animations did not include accompanying text. However, the effectiveness of an animation depends largely on its representational role (i.e., the

animation is related to the topic to be learned and is not intended to be decorative) and on an adequate level of realism. Animations seem to be most effective for the acquisition of procedural-motor knowledge, but also for declarative knowledge and problem-solving (Höffler & Leutner, 2007). In addition, a meta-analytic review of the influence of learners' spatial ability on performance shows a moderate advantage for learners with high levels of spatial ability when working with visualizations (Höffler, 2010).

Visual attention to graphs (in economic education)

Research on the effectiveness of multimedia learning environments using eye tracking is growing (Alemdag & Cagiltay, 2018), and unique contributions on perceptual processing can be expected from research on graphs in multimedia learning using eye tracking (Mayer, 2010). In STEM, graph comprehension is a frequently studied area of research, with visual attention often analyzed in eye-tracking studies (Klein et al., 2019; Tsai et al., 2011).

A systematic review of differences between expert and novice visual processing of graphs during problem solving and learning (Ruf et al., 2023)¹³ showed that the most frequently reported eye-tracking measures were dwell time, fixation duration, fixation count, transitions (between AOI), and saccades. Most studies report only a single measure and refer to line graphs in STEM. Overall, experts pay more attention to relevant information and less to irrelevant information, consistent with the information-reduction hypothesis (Haider & Frensch, 1999). In addition to working memory, and prior knowledge assessments, studies often use performance indicators to determine expertise.

Ruf et al. (2023) identified four studies on graphs in the domain of economics using eye-tracking data (Brückner et al., 2020; Peebles & Cheng, 2003; Richter et al., 2021; Vila & Gomez, 2016). To the best of our knowledge, no further studies exist. Brückner et al. (2020) investigated graph comprehension of line graphs in different contexts for physics and economics students over the course of a semester (14 weeks). The results indicate domain-specific, contextual, and conceptual differences in the development of graph comprehension over time, as well as successful transferability across contexts and concepts. However, no differences in total dwell time were found between domain experts (economics students) and non-experts (physics students). Peebles and Cheng (2003) analyzed the complex interaction behavior of people with visual displays and graphical user interfaces in an economic line graph reasoning task. They compared eye movement data with an optimal scan path and show that the paths assumed in the task

¹³ Ruf et al. (2023) point out that the definition of expertise was not consistent across studies.

analysis approximate the sequences of saccades made by individuals. Additionally, their cognitive architecture models replicate response latencies and scan paths using an adapted version based on Adaptive Control of Thought-Rational (ACT-R). Richter et al. (2021) replicated the effect of prior knowledge on learning supported by visual aids. The use of instructional aids in learning materials benefits novices but does not support, and may even hinder, the knowledge acquisition of experts (expertise reversal effect; Kalyuga, (2013). As expected, an effect was found, which could be explained by the students' level of prior knowledge and the time to first fixation on the graphs. Experts (high prior knowledge) looked at graphs later than students with low prior knowledge. Vila and Gomez (2016) analyzed differences in visualization strategies between subjects when answering a question related to an economic bar graph. The main difference in visualization strategy was that experts (high performers) spent more time on relevant AOIs compared to low performers. Overall, these findings underscore the complexity of visual information processing and the important role that prior knowledge and expertise play in how individuals interact with and interpret graphical data.

6.5 The present study

We made use of a web-based learning environment designed as a self-learning tool for learners of economics. For the purpose of this study, we slightly modified the environment by omitting the instructional material, as the participants already had prior knowledge of economics. They had to rely solely on the static and dynamic, interactive graphs to complete the tasks. In principle, the learning environment also allows to offer content for beginners, but this was not intended in this study due to the prior knowledge of the learners. Hence, participants received two sets of similar learning tasks on the general topic of supply and demand that contained both a static and a dynamic graph of the supply-and-demand model. Students were prompted to use these models to answer a series of single-choice items related to the tasks. For this study, we use performance data on learning tasks implemented in the learning environment: one comparatively easy task set on supply and demand shocks and one more complex task set on state interventions on the housing market. To clarify this distinction, we refer to them as simple and difficult task sets based on solution rates. Both task sets followed the same sequence. First, students were presented with a scenario description and used a static graph of a supply-and-demand model from which they were asked to derive effects on market parameters. Students' answers to the single-choice items were logged in. Subsequently, they received the same problem presented with a dynamic graph of the supply and demand model. Figure 6-1 illustrates a dynamic graph for the learning task on a state intervention on the housing market (difficult

task). Students were able to change the slope of the supply and demand curves, as well as manipulate the maximum price, to answer several tasks about the impact on supply and demand, the welfare of producers, consumers, and total welfare. Using the reset button, learners could return to the initial situation at any time. Hence, the analysis relies on two different tasks with two different visualization forms each. Detailed description for both task sets, including an example item, can be found in the Appendix. Followed by several single choice questions, for example: ‘The sum of producer and consumer surplus... increases / remains unchanged / decreases’.

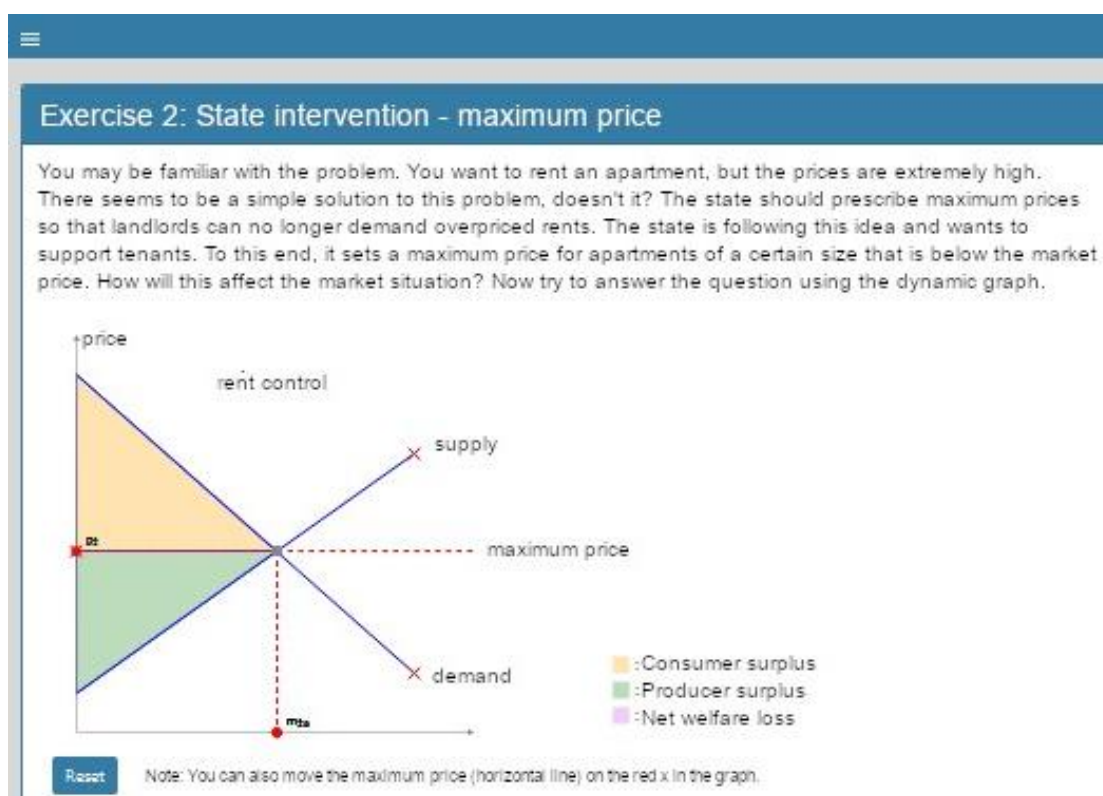


Figure 6-1: Learning task with a dynamic, interactive graph that allows manipulation to answer several single-choice tasks (not in the picture). Note. Students could change the slope of the curves and set a maximum price (drag-and-drop at the red crosses) to see changes in the market and welfare.

Following Ruf et al. (2023), three of the most prominent eye-tracking measures for graph analysis were examined in this study: dwell time (in sec), average fixation duration (in ms), and fixation count. We refer to Holmqvist and Andersson's functional operational taxonomy for eye-tracking measures (Holmqvist et al., 2011; 2017). A dwell is defined as a visit to an AOI for entry and exit, and total dwell time is the sum of dwell times (Holmqvist & Andersson, 2017, pp. 535–538). Total dwell time indicates the overall allocation of attention and may be influenced by interest and novelty but may also indicate uncertainty or difficulty in extracting information. Average fixation duration can be defined as the time the eye remains still

(Holmqvist & Andersson, 2017, pp. 526–532). Fixations are determined by the fixation detection algorithm used. Longer fixations may indicate deeper processing of an AOI, while shorter fixations may indicate less processing difficulty for experts (Gegenfurtner et al., 2011). In addition, we measure the total fixation duration as the sum of all fixation durations. The total fixation duration is expected to determine the overall amount of visual attention allocated during learning (Scheiter & Eitel, 2017). A systematic review found a generally positive link between the amount of attention allocated to relevant content and learning performance (Alemdag & Cagiltay, 2018). Last, fixation counts are the number of fixations counted in a given AOI (Holmqvist & Andersson, 2017, pp. 560–563). A higher number of fixations might indicate the importance of a stimulus, and more experienced subjects tend to have fewer fixations.

Hypotheses

In line with the theoretical background (Mayer, 2021), previous research on the learning effectiveness of dynamic visualizations (Höffler & Leutner, 2007), and eye-tracking research on graphs (Ruf et al., 2023) in the field of STEM (e.g., Jiang et al., 2019; Tsai et al., 2011) and economics (e.g., Brückner et al., 2020), we expect differences in performance and visual attention depending on the stimuli visualization (static vs. dynamic, interactive graph). Thus, we formulate the following hypotheses:

- H1: Students' performance in solving economic learning tasks is higher with dynamic graphs than with static graphs.
- H2: Students' visual attention differs between dynamic and static graphs concerning eye-tracking measures (dwell time, fixation duration, number of fixations).
- H3: Total fixation duration on graphs predicts students' task performance.

Hypothesis H3 is examined for one selected eye-tracking measure only (fixation duration) due to high dependencies between the eye-tracking measures and resulting multicollinearity concerns. We selected fixation duration based on its importance in indicating the overall amount of visual attention while working in the learning environment (Scheiter & Eitel, 2017).

6.6 Methodology

Participants

We follow the reporting practices recommended by Carter and Luke (2020) as best practice for reporting on eye movement research. To test the hypotheses, we invited students in economics and business education to participate in an eye-tracking study under laboratory conditions. As

part of the curriculum at the study program where this study was conducted, students are required to participate as subjects in empirical research to gain insight into different research methods from a participant's perspective. Opportunities to fulfil these duties are offered regularly throughout the program, and students can choose which study to participate in. A total of 33 students (18 male, 15 female, 17 Bachelor's and 16 Master's students) participated in this study. All participants were previously enrolled in the 'Introduction to Economics' course. All participants had normal or corrected-to-normal visual acuity and no history of eye disease. One participant was excluded due to poor data quality and another due to outlier detection on the eye-tracking data, leaving $n = 31$ participants.

Apparatus

The experiment was conducted using a Gazepoint GP3 HD remote eye-tracking device at a sampling rate of 150 Hz with a viewing angle of $\sim 1^\circ$ visual accuracy. Precision information is not provided by the manufacturer, but research suggests that the eye-tracking device is comparable to competing systems (Cuve et al., 2022). This precision is sufficient for the purpose of our study. Both eyes were tracked. The lab includes an external monitor (23-inch, 1920x1080 resolution) and an external mouse and keyboard in a dual-screen setup. We did not use a chin/head rest because the head movement range of the device covers most of the natural head movement behavior (35 cm horizontal x 22 cm vertical). The size of the stimuli was approximately the same in the dynamic learning environment. Participants were seated at a viewing distance of 65 cm from the monitor. iMotions (2023) was used for stimulus presentation and data collection.

Instruments

Prior knowledge was measured using seven pertinent items from the Test of Economic Literacy IV (Walstad et al., 2013) in the adapted German version (for further information see Happ, Förster, et al., 2016)¹⁴. In addition, attitude was measured by reframing a five-item subscale on general attitude (towards ICT) towards economic models (Scherer et al., 2018). An example item is 'I am interested in economic models.' ($\alpha = .75$). Finally, we calculated task performance as a dependent variable. Successful completion of a single choice question is rewarded with one point, which is added to an individual's performance rating and serves as a quantifiable measure of participants' performance and proficiency in economics, with a maximum total of 22 points possible.

¹⁴ WiWiKom-Test. Copyright © 2014 JGU Mainz, FB 03, Wirtschaftspädagogik I, Mainz. All rights reserved. For more information visit http://www.wipaed.uni-mainz.de/ls/1085_ENG_HTML.php

Procedure

On average, the experimental procedure took about 30 minutes per participant. Figure 6-2 illustrates the procedure. Prior to the experiment, all participants were given a brief introduction to the procedure and some pre-experimental explanations regarding eye-tracking. Participants then voluntarily signed the informed consent form. Afterwards, eye calibration was performed using a nine-point calibration procedure. Calibration was repeated if the quality of the calibration led to a poor result. If calibration failed more than twice, participants were allowed to complete the study but were excluded from the data analysis and not included in this sample. Participants opened the web-based learning environment and were introduced to the learning environment itself, including the handling of dynamic graphs, via video instruction. They then answered the questionnaires and completed a series of economic learning tasks in the learning environment. Finally, the experimenter thanked the participants for their participation and asked them briefly about their impressions.

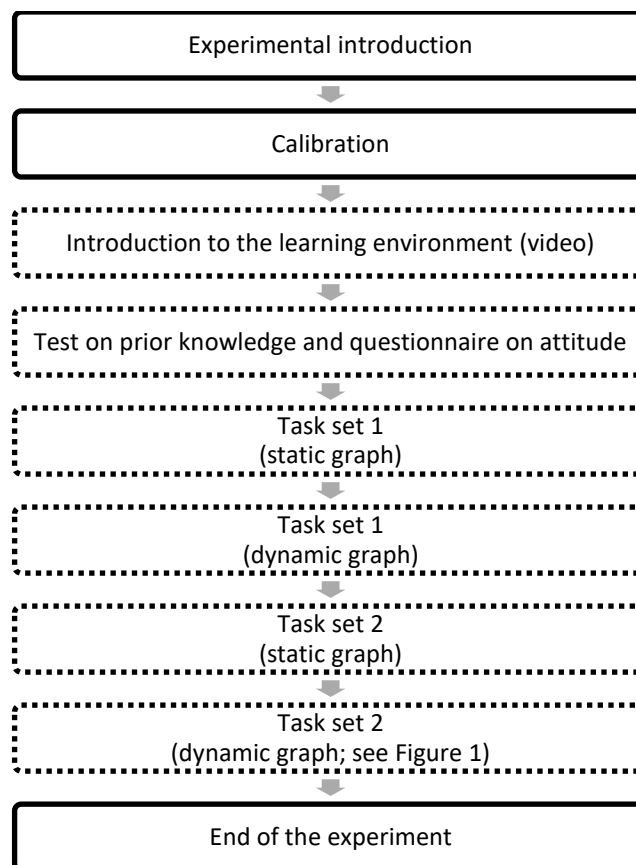


Figure 6-2: The experimental procedure. Note. Activities in the learning environment are highlighted in dashed lines.

Data analysis

Raw eye-tracking data were preprocessed using the I-VT (velocity-threshold identification) algorithm (Olsen & Matos, 2012), with a standard velocity threshold setting of 30°/sec. We further cleaned and processed our data in R (Version 4.3.2; R Core Team, 2023) and R Studio (Posit Team, 2023) using ‘tidyverse’ packages (Wickham et al., 2019) for data handling (wrangling, manipulating, and melting) such as ‘dplyr’ (Wickham et al., 2023). In addition, we used the ‘bestNormalize’ package (Peterson & Cavanaugh, 2020) to transform and standardize non-normally distributed eye-tracking variables, analogous to Cloude et al. (2020). Similarly, we eliminated one participant using Grubb's test (1969) due to significant outlying observations in the eye-tracking data.

To examine the hypothesis, whether students performed better on tasks with static graphs than on tasks with dynamic, interactive graphs (H1), we calculated two performance scores, one for tasks with static graphs and one for tasks with dynamic graphs. Performance scores represent the sum of correctly solved items. We conducted a one-tailed paired *t*-test to compare performance between tasks with static graphs and dynamic graphs. To examine whether students' visual attention differed between dynamic and static graphs on eye-tracking measures (H2), repeated measures ANOVA were conducted for each eye-tracking measures (dwell time, fixation duration, number of fixations) on static and dynamic graphs and a post-hoc test on the task level was conducted. Finally, fixation duration per task was used as a predictor in multiple linear regression analyses as a measure of overall visual attention to the graphs to predict mean student performance on dynamic or static tasks (H3). Assumptions were checked for and variance inflation factors were acceptable. Additionally, Durbin-Watson tests have shown no critical violations of independence. Prior knowledge and attitudes towards economic graphs were included as control variables, as previous studies have shown that these variables (Cohn et al., 2004; Hill & Stegner, 2003; Kalyuga, 2013) can have an impact on learning with graphs. No major correlations were found between the controls (prior knowledge, attitude) and the total fixation duration as additional predictor.

6.7 Results

Descriptives

Table 6-1 summarizes individual scores per exercise to examine differences in performance across all tasks and graphs. Performance appears to be evenly distributed across participants. For the first task set (external shocks), there appears to be no difference at all between students'

performance when working with static graphs and their performance when using dynamic, interactive graphs. On average, participants successfully completed around 80% of the external shocks related items (easier task) for both types of graphs. For the second and more difficult set of tasks (government intervention) in terms of solution rates, there seems to be a small performance advantage for dynamic graphs. On average, participants solved 60% of the single-choice tasks supported by the static graph and 70% of the tasks supported by the dynamic graph.

Table 6-1: Performance Scores of Participants.

	<i>n</i>	Min	Max	<i>M</i>	<i>SD</i>	<i>SE</i>
Static Graph	31	1	6	4.81	1.38	0.25
Dynamic Graph	31	1	6	4.84	1.37	0.25
Static Graph	31	1	5	3.00	1.32	0.24
Dynamic Graph	31	1	5	3.52	1.46	0.26

Note. The Min and Max values refer to the lowest and highest possible values as well as to the actual lowest and highest performance.

Hypothesis testing

We tested the differences in performance between the two sets of tasks (H1). No difference in performance was found in the first set of tasks ($t(30) = 0.13, p = 0.45$). The results for the second task set indicate a difference in performance between static and dynamic, interactive graphs. Performance on tasks supported by dynamic graphs ($M = 3.52, SD = 1.46$) was significantly higher than performance on tasks supported by static graphs ($M = 3.00, SD = 1.32$), $t(30) = 2.89, p < .01, d = 0.52$. According to Cohen (1988), this can be interpreted as a medium sized effect.

Regarding Hypothesis 2, whether visual attention differs between dynamic and static graphs concerning eye-tracking measures, the results indicate no significant differences between static and dynamic graphs for dwell time and number of fixations (Table 6-2). However, fixation duration was significantly different between static and dynamic graphs ($p < .001, \eta_p^2 = 0.18$), and partial eta square indicates a large effect size (Cohen, 1988).

Table 6-2: Repeated measurement ANOVA of eye-movement measures.

Measures		Type	Static Graph	Dynamic Graph	<i>F</i>	<i>p</i>	η_p^2
Dwell Time (sec)		M	32.5	30.1	0.45	.508	-
		SD	13.0	14.8			
Fixation (ms)	Duration	M	106	114	6.42	<.001 (***)	0.18
		SD	11.3	12.9			
Fixation Count		M	83.5	81.8	0.03	.867	-
		SD	41.2	47.2			

Note. Sphericity correction method: Greenhouse-Geisser

To gain further insights into differences regarding fixation duration, we additionally conducted a post-hoc analysis at the tasks level. The results (Table 6-3) indicated that three of four pairwise differences between static and dynamic, interactive graphs were statistically significant ($p < .05$). However, the comparison between the easy task with static graphs and the complex task with dynamic graphs was not significant at the 5% level. In line with our hypothesis, no significant differences were found when comparing the two static graphs to each other or the two dynamic graphs to each other.

Table 6-3: Post hoc analysis for Fixation Duration (ms).

Metrics		Comparison	<i>Estimate</i>	<i>SE</i>	<i>t</i>	<i>p</i>
Main Effect:						
		SG – DG	-7.7	1.78	-4.318	<.001 (***)
Task Level Contrasts:						
Fixation (ms)	Duration	SG 1 – SG 2	1.1	2.64	0.410	.9763
		SG 1 – DG 1	-8.7	2.48	-3.493	<.01 (**)
		SG 1 – DG 2	-5.6	2.25	-2.506	.0793
		SG 2 – DG 1	-9.8	3.14	-3.108	<.05 (*)
		SG 2 – DG 2	-6.7	2.31	-2.907	<.05 (*)
		DG 1 – DG 2	3.0	2.60	1.165	.6531

Note. SG = Static Graphs, DG = Dynamic Graph, SG 1 = Static Graph 1, SG 2 = Static Graph 2, DG 1 = Dynamic Graph 1, DG 2 = Dynamic Graph 2, p-value adjustment: tukey method

Finally, Table 6-4 and Table 6-5 show the findings for the multiple regression analysis for static and dynamic, interactive diagrams respectively. Descriptive statistics for the final models can be found in the appendix (Table 6-6 and 6-7). In model 1a ($F(2,59) = 6.87, p < .01$), prior knowledge ($\beta = .37, p < .01$) significantly predicted mean scores for the tasks supported by static graphics, while attitude towards economic models ($\beta = .18, p = .13$) did not. The inclusion of total fixation duration in model 1b ($F(3,58) = 4.60, p < .01$) did not increase the total variance

explained; it was a non-significant predictor ($\beta = .06, p = .62$) for mean scores on the static graph tasks. In model 2a ($F(2,59) = 9.91, p < .001$), prior knowledge ($\beta = .39, p < .01$) and attitude ($\beta = .26, p < .05$) significantly predicts mean scores on the dynamic, interactive graph tasks. Including total fixation duration in model 2b ($F(3,58) = 10.63, p < .001$) significantly predicts ($\beta = .33, p < .01$) mean scores on the dynamic graph tasks (see Fig. 6-3). The final model explained 36% of the variance in the mean score for dynamic graphs (adjusted $R^2 = 32\%$).

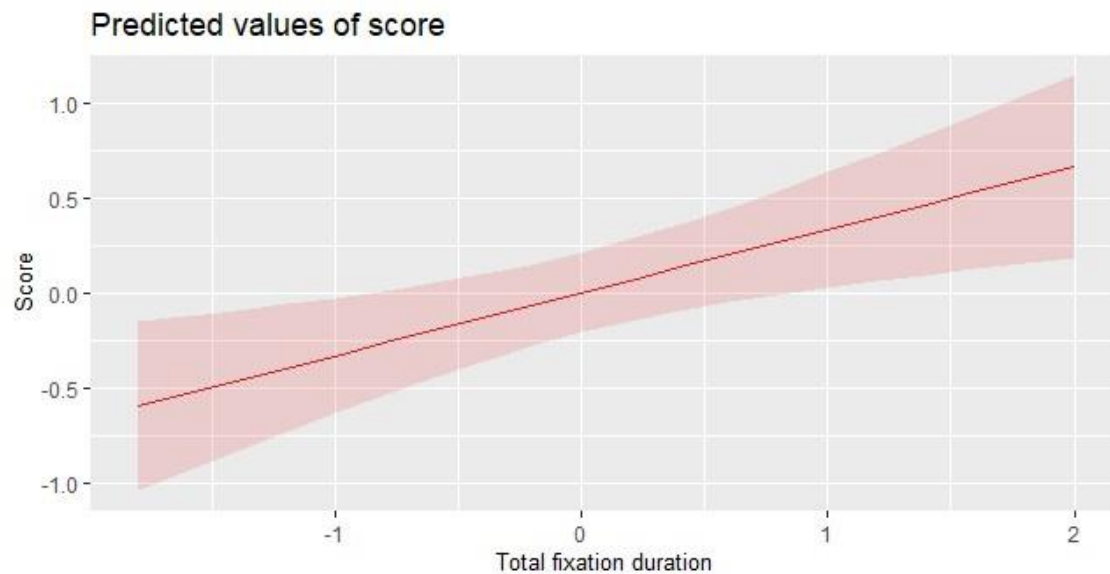


Figure 6-3: The predicted mean score for dynamic, interactive graphics exercises by average fixation duration (both standardized).

Table 6-4: Regression results using mean performance as the dependent variable.

<i>Predictors</i>	Model 1a (static)				Model 1b (static)			
	<i>Estimate</i>	<i>SE</i>	β	<i>t</i>	<i>Estimate</i>	<i>SE</i>	β	<i>t</i>
(Intercept)	0.49	1.22	-	0.40	0.33	1.26	-	0.27
Prior Knowledge	0.32**	0.10	0.37	3.10	0.33**	0.11	0.39	3.09
Attitude	0.59	0.39	0.18	1.52	0.62	0.40	0.19	1.56
Total Fixation Duration ^a	-	-	-	-	0.10	0.20	0.06	0.50
Adjusted R^2	-	.16	-	-	-	.15	-	-
F	-	6.87**	-	-	-	4.60**	-	-

Note. ^a standardized coefficient. A significant Estimate indicates the β -weight is also significant. * Indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$.

Table 6-5: Regression results using mean performance as the dependent variable.

<i>Predictors</i>	Model 2a (dynamic)				Model 2b (dynamic)			
	<i>Estimate</i>	<i>SE</i>	β	<i>t</i>	<i>Estimate</i>	<i>SE</i>	β	<i>t</i>
(Intercept)	0.06	1.12	-	0.05	-0.74	1.08	-	-0.68
Prior Knowledge	0.32**	0.09	0.39	3.42	0.33***	0.09	0.39	3.67
Attitude	0.82*	0.36	0.26	2.27	1.09**	0.35	0.34	3.11
Total Fixation Duration ^a	-	-	-	-	0.51**	0.17	0.33	3.05
Adjusted R^2	-	.23	-	-	-	.32	-	-
F	-	9.91***	-	-	-	10.63***	-	-

Note. ^a standardized coefficient. A significant *Estimate* indicates the β -weight is also significant. * Indicates $p < .05$. ** indicates $p < .01$. *** indicates $p < .001$.

6.8 Limitations

This study has several limitations. While there are some indicators of prior knowledge for our students, it must be clear that no ceiling effect was found for the participants performance and that the participants of the study are learners rather than experts. Most of them do not take economics classes beyond the mandatory ones in the curriculum. In addition, the tasks were not randomized, but each set of tasks had the same sequence: static graphs were presented first and dynamic graphs second. Consequently, this study can yield information about the *additional* value of dynamic graphs on students' performance. When interpreting the results, one has to keep in mind that the task as well as the single-choice items students should answer were the same across each of the two sequences. The element that was added in the second step was the possibility to interactively engage with a dynamic graph. Moreover, tasks were built in from easier to more difficult ones and subjective ratings were chosen for the difficulty of the tasks, but the solution rates support this differentiation. Additionally, regarding the metrics obtained, Negi and Mitra (2020) point out that the arithmetic mean of fixation durations, which is one of the most commonly reported metric for eye-tracking research with graphs, may not be the optimal fixation duration metric. They suggest models of fixation durations where the number of fixations over different temporal ranges are used. Furthermore, Kosel et al. (2023) show in an eye-tracking study of professional vision that the chosen velocity threshold parameter can be altered and, in the worst case, bias the results of an eye-tracking study. However, as reported, we stick to the standard values for identifying fixations. In addition, while we follow the standard recommendations for fixation classification in this study, classification rules play an important role in the merging and selection of fixation candidates (Hooge et al., 2022). Other eye-tracking measures also did not show significant within-subject differences, so we maintain fixation duration as the metric of choice for further analysis. This may be attributed to the low-stakes test environment. In addition, this study did not focus on transfer performance, i.e., we did not include a subsequent transfer test of performance, but rather focused on retention and solving the single-choice items. A delayed post-test would be needed to see if long-term learning effects occurred. The students' eye movement behavior may be explained by the low-stakes test environment, which they viewed as a chore for program credit, leading to a low level of engagement.

6.9 Discussion

Previous research suggests that there is an urgent need for economic education (Walstad & Rebeck, 2001) due to a lack of economic knowledge as a prerequisite for making well-informed

life decisions. The described learning environment is supported by static and dynamic, interactive graphics. Thus, this paper investigates performance and the visual attention when using static and dynamic, interactive graphs of price-quantity graphs to solve economic tasks for economic and business education students. This paper can, therefore, provide insights regarding the *additional* value of dynamic graphs compared to static graphs with respect to performance on economic learning tasks. Using eye tracking as an unobtrusive measure of students' learning behavior, we observe and analyze students' visual attention as they solve learning tasks in a multimedia learning environment.

In a first step, performance differences were analyzed and students' overall performance in solving economic learning tasks was higher for exercises with dynamic, interactive graphs compared to static graphs. However, the effect was only significant for the more difficult task set. Therefore, while hypothesis 1 can be partly supported, it must be emphasized that these findings need to be further investigated. However, these results are consistent with previous research in this area. Research in mathematics education has shown that dynamic graphs help students to reason better when solving mathematical problems (Lassak, 2009), which could explain the significant findings for the more difficult tasks, as these require also more reasoning to solve the learning tasks. Overall, previous research suggests these moderate advantages of dynamic visualizations over static visualizations in learning (Höffler & Leutner, 2007).

Next, differences in the visual attention of economic education students while solving economic learning tasks were measured using eye tracking. The most prominent eye-tracking measures (dwell time, fixation duration, number of fixations) for analyzing graphs (Ruf et al., 2023) were examined, and significant differences in average fixation durations for visual attention to dynamic and static graphs were found. Subjects fixated on dynamic graphs on average significantly longer than on static graphs, and task level comparisons show that this is the case for most of the comparison between static and dynamic graphs. Based on these results, hypothesis H2 can be supported. In general, participants did not spend much time fixating on the graph while solving the tasks. The amount of time spent on the actual graphs is rather low and especially fixation counts, and average fixation durations indicates a fast visual processing of the presented material. In the literature, average fixation durations between 90 and 150 ms, as observed in this study, are sometimes referred to as express fixations (Holmqvist & Andersson, 2017) or ambient fixations (Negi & Mitra, 2020), which, according to the literature, are usually not associated with deep cognitive processing, but rather indicate simple cognitive processes on visual targets. Longer fixations (150 - 900 ms) allow for more complex cognitive operations, as we would have expected for these tasks. For example, average fixation durations for visual

search tasks are expected to be in the range of 180 - 275 ms, while those for reading are expected to be 225 - 250 ms (Rayner & Castelano, 2007).

Several factors could explain the rather short average fixation durations observed in this study. First, the familiarity with the material and task type may have contributed to this behavior. Many of the participants in this study had a certain amount of prior knowledge about the subject. A meta-analysis of expertise effects on visual comprehension by Gegenfurtner et al. (2011) confirmed that experts had shorter fixation durations, more fixations to relevant areas, and fewer fixations to irrelevant areas than novices. Experts also showed selective attention through parafoveal processing. In contrast, explicit findings on expertise effects on visual attention to graphs by Ruf et al. (2023) report mixed results for fixation durations. For instance, Ahmed et al. (2021) report longer fixation durations on average, but experts spend less time on the graph than non-experts. Another reason can be related to the effect of visualization features that influence visual attention. For dynamic visualizations, experts show more fixations of shorter duration compared to novices who tend to show the same behavior for static visualizations (Gegenfurtner et al., 2011).

Finally, total fixation duration was included as a predictor in multiple regression to see if the metric can predict performance. Total fixation durations predicted students' mean performance on tasks supported by dynamic, interactive graphs, while this was not the case for static graphs. Total fixation duration can explain a substantial fraction of the variation in performance, indicating its relevance in examining learning processes. However, we did not expect to find a significant prediction only for performance on tasks supported by dynamic graphics. Thus, hypothesis H3 can only be partially supported. This could be due to the fact that the students are familiar with static diagrams from relevant lectures.

Additionally, research from more complex learning environments such as serious games suggests that combining different data streams, such as log file data combined with gaze data, can explain a substantial proportion of the variance in performance (Taub et al., 2017), and that machine learning models that implement multimodal data for selected features, such as gaze behavior and facial expressions, can be helpful to predict different performance groups (e.g., Emerson et al., 2020). Future research should provide more meaningful incentives for participation or target vocational students or university students from non-economic and business disciplines unfamiliar with the material. These participants are likely to approach the task with more deliberation, potentially revealing more pronounced differences in visual attention.

6.10 References

- Ahmed, A., Hurwitz, D., Gestson, S., & Brown, S. (2021). Differences between Professionals and Students in Their Visual Attention on Multiple Representation Types While Solving an Open-Ended Engineering Design Problem. *Journal of Civil Engineering Education*, *147*(3), 04021005. [https://doi.org/10.1061/\(ASCE\)EI.2643-9115.00000044](https://doi.org/10.1061/(ASCE)EI.2643-9115.00000044)
- Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, *125*, 413–428. <https://doi.org/10.1016/j.compedu.2018.06.023>
- Baddeley, A. D. (1997). *Human memory: Theory and practice*. psychology press.
- Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education*, *101*, 150–167. <https://doi.org/10.1016/j.compedu.2016.06.005>
- Brückner, S., Zlatkin-Troitschanskaia, O., Küchemann, S., Klein, P., & Kuhn, J. (2020). Changes in Students' Understanding of and Visual Attention on Digitally Represented Graphs Across Two Domains in Higher Education: A Postreplication Study. *Frontiers in Psychology*, *11*, 2090. <https://doi.org/10.3389/fpsyg.2020.02090>
- Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, *155*, 49–62.
- Chaudhury, P. (2023). Asynchronous learning design—Lessons for the post-pandemic world of higher education. *The Journal of Economic Education*, *54*(2), 214–223. <https://doi.org/10.1080/00220485.2023.2174233>
- Cloude, E. B., Dever, D. A., Wiedbusch, M. D., & Azevedo, R. (2020). Quantifying Scientific Thinking Using Multichannel Data With Crystal Island: Implications for Individualized Game-Learning Analytics. *Frontiers in Education*, *5*, 572546. <https://doi.org/10.3389/feduc.2020.572546>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Academic press.
- Cohn, E., & Cohn, S. (1994). Graphs and learning in principles of economics. *The American Economic Review*, *84*(2), 197–200.
- Cohn, E., Cohn, S., Balch, D. C., & Bradley, J. (2001). Do Graphs Promote Learning in Principles of Economics? *The Journal of Economic Education*, *32*(4), 299–310. <https://doi.org/10.1080/00220480109596110>
- Cohn, E., Cohn, S., Balch, D. C., & Bradley, J. (2004). The Relation between Student Attitudes toward Graphs and Performance in Economics. *The American Economist*, *48*(2), 41–52. <https://doi.org/10.1177/056943450404800203>
- Cuve, H. C., Stojanov, J., Roberts-Gaal, X., Catmur, C., & Bird, G. (2022). Validation of Gaze-point low-cost eye-tracking and psychophysiology bundle. *Behavior Research Methods*, *54*(2), 1027–1049. <https://doi.org/10.3758/s13428-021-01654-x>
- Duchowski, A. T. (2017). *Eye Tracking Methodology*. Springer. <https://doi.org/10.1007/978-3-319-57883-5>
- Emerson, A., Cloude, E. B., Azevedo, R., & Lester, J. (2020). Multimodal learning analytics for game-based learning. *British Journal of Educational Technology*, *51*(5), 1505–1526. <https://doi.org/10.1111/bjet.12992>
- Findeisen, S., Messner, J., & Guggemos, J. (2022). Dynamische, interaktive Visualisierungen – Erkenntnisse eines systematischen Reviews und Vorstellung einer digitalen Lernumgebung für die Domäne Wirtschaft [Dynamic, interactive visualizations—Findings of a

- systematic review and presentation of a digital learning environment for the business domain]. In K. Kögler, U. Weyland, & H.-H. Kremer (Eds.), *Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2022* (pp. 219–235). Barbara Budrich. <https://doi.org/10.25656/01:26524>
- Friebel-Piechotta, S. (2022). Economic models in German textbooks. *JSSE - Journal of Social Science Education*, 21(2), 82–113. <https://doi.org/10.11576/JSSE-4854>
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains. *Educational Psychology Review*, 23(4), 523–552. <https://doi.org/10.1007/s10648-011-9174-7>
- Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics*, 11(1), 1–21. <https://doi.org/10.1080/00401706.1969.10490657>
- Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(1), 172–190. <https://doi.org/10.1037/0278-7393.25.1.172>
- Happ, R., Förster, M., Zlatkin-Troitschanskaia, O., & Carstensen, V. (2016). Assessing the previous economic knowledge of beginning students in Germany: Implications for teaching economics in basic courses. *Citizenship, Social and Economics Education*, 15(1), 45–57. <https://doi.org/10.1177/2047173416646597>
- Happ, R., Zlatkin-Troitschanskaia, O., & Schmidt, S. (2016). An analysis of economic learning among undergraduates in introductory economics courses in Germany. *The Journal of Economic Education*, 47(4), 300–310. <https://doi.org/10.1080/00220485.2016.1213686>
- Hill, C. D., & Stegner, T. (2003). Which Students Benefit from Graphs in a Principles of Economics Class? *The American Economist*, 47(2), 69–77. <https://doi.org/10.1177/056943450304700206>
- Höffler, T. N. (2010). Spatial Ability: Its Influence on Learning with Visualizations—a Meta-Analytic Review. *Educational Psychology Review*, 22(3), 245–269. <https://doi.org/10.1007/s10648-010-9126-7>
- Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17(6), 722–738. <https://doi.org/10.1016/j.learninstruc.2007.09.013>
- Holmqvist, K., & Andersson, R. (2017). *Eye tracking: A comprehensive guide to methods, paradigms, and measures* (2nd edition). Lund Eye-Tracking Research Institute.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- Hooge, I. T. C., Niehorster, D. C., Nyström, M., Andersson, R., & Hessels, R. S. (2022). Fixation classification: How to merge and select fixation candidates. *Behavior Research Methods*, 54(6), 2765–2776. <https://doi.org/10.3758/s13428-021-01723-1>
- Hoyt, G. M., & McGoldrick, K. (2019). 50 years of economic instruction in the Journal of Economic Education. *The Journal of Economic Education*, 50(2), 168–195. <https://doi.org/10.1080/00220485.2019.1582388>
- iMotions (9.3). (2023). [Computer software]. iMotions A/S.
- Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in Educational Science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research*, 10(1), 1–18. <https://doi.org/10.16910/JEMR.10.1.3>

- Jiang, J.-Y., Guo, F., Chen, J.-H., Tian, X.-H., & Lv, W. (2019). Applying eye-tracking technology to measure interactive experience toward the navigation interface of mobile games considering different visual attention mechanisms. *Applied Sciences*, 9(16), 3242.
- Kalyuga, S. (2013). Effects of Learner Prior Knowledge and Working Memory Limitations on Multimedia Learning. *Procedia - Social and Behavioral Sciences*, 83, 25–29. <https://doi.org/10.1016/j.sbspro.2013.06.005>
- Kalyuga, S., & Sweller, J. (2014). The Redundancy Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 247–262). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.013>
- Klein, P., Küchemann, S., Brückner, S., Zlatkin-Troitschanskaia, O., & Kuhn, J. (2019). Student Understanding of Graph Slope and Area under a Curve: A Replication Study Comparing First-Year Physics and Economics Students. *Physical Review Physics Education Research*, 15(2), 020116. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020116>
- Kosel, C., Voggenreiter, A., Pfeffer, J., & Seidel, T. (2023). Measuring Teachers' Visual Expertise Using the Gaze Relational Index Based on Real-World Eyetracking Data and Varying Velocity Thresholds. *Journal of Expertise*, 6(2), 267–281.
- Kourilsky, M., & Wittrock, M. C. (1987). Verbal and graphical strategies in the teaching of economics. *Teaching and Teacher Education*, 3(1), 1–12. [https://doi.org/10.1016/0742-051X\(87\)90030-8](https://doi.org/10.1016/0742-051X(87)90030-8)
- Lassak, M. (2009). Using dynamic graphs to reveal student reasoning. *International Journal of Mathematical Education in Science and Technology*, 40(5), 690–696. <https://doi.org/10.1080/00207390902759642>
- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction*, 20(2), 167–171. <https://doi.org/10.1016/j.learninstruc.2009.02.012>
- Mayer, R. E. (2021). *Multimedia Learning* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781316941355>
- Nakayama, M., & Santiago, R. (2012). Learner Characteristics and Online Learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 1745–1747). Springer. https://doi.org/10.1007/978-1-4419-1428-6_1858
- Negi, S., & Mitra, R. (2020). Fixation duration and the learning process: An eye tracking study with subtitled videos. *Journal of Eye Movement Research*, 13(6), 1–15. <https://doi.org/10.16910/jemr.13.6.1>
- Olsen, A., & Matos, R. (2012). Identifying parameter values for an I-VT fixation filter suitable for handling data sampled with various sampling frequencies. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 317–320). Association for Computing Machinery. <https://doi.org/10.1145/2168556.2168625>
- Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford university press.
- Peebles, D., & Cheng, P. C.-H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors*, 45(1), 28–46.
- Peterson, R. A., & Cavanaugh, J. E. (2020). Ordered quantile normalization: A semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*, 47(13–15), 2312–2327. <https://doi.org/10.1080/02664763.2019.1630372>
- Ploetzner, R., & Lowe, R. (2004). Dynamic visualisations and learning. *Learning and Instruction*, 14(3), 235–240. <https://doi.org/10.1016/j.learninstruc.2004.06.001>

- Posit Team. (2023). *RStudio: Integrated Development Environment for R* [Computer software]. Posit Software, PBC. <http://www.posit.co/>
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rayner, K., & Castelano, M. (2007). Eye Movements during Reading, Scene Perception, Visual Search, and While Looking at Print Advertisements. In M. Wedel & R. Pieters (Eds.), *Visual Marketing*. Psychology Press. <https://doi.org/10.4324/9780203809617>
- Richter, J., Wehrle, A., & Scheiter, K. (2021). How the poor get richer: Signaling guides attention and fosters learning from text-graph combinations for students with low, but not high prior knowledge. *Applied Cognitive Psychology*, 35(3), 632–645.
- Rieber, L. P. (1991). Animation, incidental learning, and continuing motivation. *Journal of Educational Psychology*, 83(3), 318–328. <https://doi.org/10.1037/0022-0663.83.3.318>
- Rodgers, T. L., Bolton, P., & Van Der Gryp, P. (2023). An interactive graph resource for chemical engineering teaching. *Education for Chemical Engineers*, 45, 80–89. <https://doi.org/10.1016/j.ece.2023.08.004>
- Ruf, V., Horrer, A., Berndt, M., Hofer, S. I., Fischer, F., Fischer, M. R., Zottmann, J. M., Kuhn, J., & Küchemann, S. (2023). A Literature Review Comparing Experts' and Non-Experts' Visual Processing of Graphs during Problem-Solving and Learning. *Education Sciences*, 13(2), 216. <https://doi.org/10.3390/educsci13020216>
- Scheiter, K., & Eitel, A. (2017). The Use of Eye Tracking as a Research and Instructional Tool in Multimedia Learning. In C. Was, F. Sansosti, & B. Morris (Eds.), *Eye-Tracking Technology Applications in Educational Research* (pp. 143–164). IGI Global. <https://doi.org/10.4018/978-1-5225-1005-5>
- Scherer, R., Tondeur, J., Siddiq, F., & Baran, E. (2018). The importance of attitudes toward technology for pre-service teachers' technological, pedagogical, and content knowledge: Comparing structural equation modeling approaches. *Computers in Human Behavior*, 80, 67–80. <https://doi.org/10.1016/j.chb.2017.11.003>
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. [https://doi.org/10.1016/S0959-4752\(02\)00017-8](https://doi.org/10.1016/S0959-4752(02)00017-8)
- Seel, N. M. (2012). Attitudes – Formation and Change. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 371–374). Springer. https://doi.org/10.1007/978-1-4419-1428-6_359
- Sierra, J. (2020). The importance of simulation in teaching and learning economics: The students' perspective. *Innovations in Education and Teaching International*, 57(5), 521–531. <https://doi.org/10.1080/14703297.2019.1647268>
- Sorden, S. D. (2013). The cognitive theory of multimedia learning. In B. J. Irby, G. Brown, R. Lara-Alecio, & S. Jackson (Eds.), *The handbook of educational theories*. (pp. 155–167). IAP.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>

- Taub, M., Mudrick, N. V., Azevedo, R., Millar, G. C., Rowe, J., & Lester, J. (2017). Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with Crystal Island. *Computers in Human Behavior*, *76*, 641–655. <https://doi.org/10.1016/j.chb.2017.01.038>
- Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., & Yang, F.-Y. (2011). Visual Attention for Solving Multiple-Choice Science Problem: An Eye-Tracking Analysis. *Computers & Education*, *58*(1), 375–385. <https://doi.org/10.1016/j.compedu.2011.07.012>
- Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies*, *57*(4), 247–262. <https://doi.org/10.1006/ijhc.2002.1017>
- Van Gog, T., Paas, F., Marcus, N., Ayres, P., & Sweller, J. (2009). The Mirror Neuron System and Observational Learning: Implications for the Effectiveness of Dynamic Visualizations. *Educational Psychology Review*, *21*(1), 21–30. <https://doi.org/10.1007/s10648-008-9094-3>
- Vila, J., & Gomez, Y. (2016). Extracting business information from graphs: An eye tracking experiment. *Journal of Business Research*, *69*(5), 1741–1746.
- Walstad, W. B., & Rebeck, K. (2001). Assessing the economic understanding of US High-school students. *American Economic Review*, *91*(2), 452–457.
- Walstad, W. B., Rebeck, K., & Butters, R. B. (2013). The test of economic literacy: Development and results. *The Journal of Economic Education*, *44*(3), 298–309.
- Wheat, I. D. (2009). Teaching economics as if time mattered. In J. Reardon (Ed.), *The Handbook of Pluralist Economics Education* (pp. 69–89). Routledge.
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation* [Computer software]. <https://dplyr.tidyverse.org>
- Zetland, D., Russo, C., & Yavapolkul, N. (2010). Teaching Economic Principles: Algebra, Graph or Both? *The American Economist*, *55*(1), 123–131. <https://doi.org/10.1177/056943451005500113>
- Zhuang, Q., & Liu, X. (2022). Comparison of graph and animation: An unbalanced battle over two decades. *Frontiers in Psychology*, *13*, 810557. <https://doi.org/10.3389/fpsyg.2022.810557>

7 Prompt text classifications with transformer models! An exemplary introduction to prompt-based learning with large language models

Study 5 was published in November 2022 in ‘Journal of Research on Technology in Education’. The publication is available at <https://doi.org/10.1080/15391523.2022.2142872>. Reproduced with permission from Taylor & Francis.^{15 16}

7.1 Abstract

This study investigates the potential of automated classification using prompt-based learning approaches with transformer models (large language models trained in an unsupervised manner) for a domain-specific classification task. Prompt-based learning with zero or few shots has the potential to (1) make use of artificial intelligence without sophisticated programming skills and (2) make use of artificial intelligence without fine-tuning models with large amounts of labeled training data. We apply this novel method to perform an experiment using so-called zero-shot classification as a baseline model and a few-shot approach for classification. For comparison, we also fine-tune a language model on the given classification task and conducted a second independent human rating to compare it with the given human ratings from the original study. The used dataset consists of 2,088 email responses to a domain-specific problem-solving task that were manually labeled for their professional communication style. With the novel prompt-based learning approach, we achieved a Cohen’s kappa of .40, while the fine-tuning approach yields a kappa of .59, and the new human rating achieved a kappa of .58 with the original human ratings. However, the classifications from the machine learning models have the advantage that each prediction is provided with a reliability estimate allowing us to identify responses that are difficult to score. We therefore argue that response ratings should be based on a reciprocal workflow of machine raters and human raters, where the machine rates easy to classify responses and the human raters focus and agree on the responses that are difficult to classify. Further, we believe that this new, more intuitive, prompt-based learning approach will enable more people to use artificial intelligence.

¹⁵ Acknowledgement: Prompt text classifications with transformer models! An exemplary introduction to prompt-based learning with large language models, by Christian W. Mayer, Sabrina Ludwig and Steffen Brand, published in ‘Journal of Research on Technology in Education’ © copyright # [2022], reprinted by permission of Informa UK Limited, trading as Taylor & Francis Group, <http://www.tandfonline.com>.

¹⁶ Note: The chapter headings, figures, and tables of the published version have been numbered according to the numbering of this thesis. Minor changes or deviations from the published paper publication are possible. Spelling and grammar have been adjusted to American English standards. Any misspellings and typographical errors detected in the course of a further thorough proofreading have been corrected.

Keywords: artificial intelligence in education, machine learning, natural language processing, transformer-based language models, prompt-based learning, classification

7.2 Introduction

Artificial intelligence (AI) is expected to impact education significantly. One major consequence is seen in the need to rethink the role of teachers (see Popenici & Kerr, 2017). Speech recognition, text translation, and product recommendations are just a few examples of ubiquitous AI algorithms. In the last twenty years, new AI technologies have emerged, and research on AI in Education (AIED) has evolved from teacher-centered and technology-based design research to student-centered research focused on learner profiles and individual learning analytics (Guan et al., 2020). AIED can be defined, among many other definitions, as “computing systems that are able to engage in human-like processes such as learning, adapting, synthesizing, self-correction and use of data for complex processing tasks” (Popenici & Kerr, 2017, p. 2). In this paper, we demonstrate a novel method for AIED to work with large language models without training data for a scoring task.

Scoring of open text answers is a time-consuming complex processing task that is often considered a major burden among professionals. Automated Essay Scoring (AES) can be a helpful and time-saving tool to support the assessment of student work. AES tools typically extract general natural language processing (NLP) features such as the length of a sentence, the number of capitalized words, or domain-specific features such as similarity or frequency of individual words (Attali & Burstein, 2006; Ifenthaler, 2014, for a comprehensive review see also Ifenthaler, 2022). AES research relies on supervised machine learning algorithms, such as support vector machines (Bin & Jian-Min, 2011; Yamamoto et al., 2018) or deep neural networks (Liang et al., 2018; Taghipour & Ng, 2016) that are trained on task-specific datasets including thousands of domain-specific text examples and handcrafted labels. Moreover, such AES tools are developed by experts and computer scientists for specific tasks and are pretty expensive. Therefore, AES tools are mainly used in large-scale assessments. In 2017, a new type of machine-learning model, called ‘Transformer’ (Vaswani et al., 2017), was introduced by Google Brain. Since then, transformer models have been outperforming previous state-of-the-art deep learning models on NLP tasks and have become the dominant model architecture in NLP (Brown et al., 2020; Radford et al., 2019; Vaswani et al., 2017; see also paperswithcode.com for current state-of-the-art research). Transformers are large language models trained to recognize patterns in natural language. To do so, they use a self-attention mechanism that enables a better understanding of words in context that outperforms non-transformer models in a wide variety of NLP tasks (Gillioz et al., 2020; Qiu et al., 2020). The model architecture has even

adapted to and proven superior in related AI fields such as computer vision (Han et al., 2022). The advantages of transformers resulted in a significant shift towards using pre-trained language models (LM). These models are trained on a tremendous amount of unlabeled data (unsupervised learning). Subsequently, they are usually fine-tuned (a common practice in deep learning to adjust a model on a specific task and domain) with several thousand training examples including text and labels (supervised learning) to downstream tasks, such as domain-specific text classification or text generation. Despite the clear advantages of transformers and the recent shift to transformer models, the need for large amounts of fine-tuning data, proficient programming skills, and a deeper understanding of neural networks is essential, makes it difficult to adapt the application to practice-oriented fields such as educational research and practice.

Recently, prompt-based learning, a less technical no-code approach to ‘instruct’ large language models on a given task, has shown promising results (see, for example, Brown et al., 2020). Prompts are short pieces of texts in natural language that can include examples of the task to be completed (few-shot learning) or just an instruction without examples (zero-shot learning; Yin et al., 2019). A simple zero-shot text classification prompt might, for example, look like this:

Text: This is a fantastic soccer game!

Sentiment (Positive, Negative):

Where ‘This is a fantastic soccer game!’ is the text that is to be classified according to its sentiment. The task of the language model is then to continue the overall prompt input with the most likely word (or often words). For the given example, the language model would, based on the millions of texts it has been trained on, estimate that the words ‘Positive’ and ‘Negative’ have the highest probabilities to follow in such a text, and since the text before was very positive and sentiment apparently plays a role, it will predict the word ‘Positive’ the highest probability¹⁷. This approach allows people to use AI even without programming skills. However, the challenge is now to design prompts in a meaningful way (prompt engineering). In this paper we demonstrate the use of prompt design for text classification and investigate its potential with a brief experiment on a domain-specific and practice-relevant classification task. More precisely, we investigate to what extent a prompt-based learning approach can yield results comparable

¹⁷ Try this example for yourself. We have created an interface with a simple demo app: <https://huggingface.co/spaces/chrismay/Sentiment-demo-app> (Please be patient until the model is loaded)

to an approach based on fine-tuning a model with a large amount of task-specific training data (Ludwig et al., 2021). To do so, we use a dataset of a domain-specific problem-solving task, which was manually labeled for their professional communication style. Word choice and sentence structure variety can be considered as one (of many) dimension-specific AES tasks (Ke & Ng, 2019). Since there is as yet little research on the potential of transformer-based language models and particularly prompt-based learning in the context of educational NLP tasks, we investigate the following questions:

- (1) To what extent does prompt-based learning with current state-of-the-art language models utilizing zero- and few-shot prompts provide reliable results for assessing politeness in an automated text scoring task?
- (2) To what extent does prompt-based learning provide comparable results to a fine-tuned language model?
- (3) What is the potential of machine ratings to improve human ratings?

Before doing so, we first consider work related to the field of text classification using prompt-based and fine-tuning approaches for language models. Second, we describe the methodological approach of prompt-based learning in more detail. Thereafter, we describe the study design and report its results. Finally, we discuss both the potential and limitations of this methodological approach.

7.3 Related Work

With the rise of, and increasing attention on, pre-trained language models (PLM) such as GPT (Generative Pre-trained Transformer; see Radford et al., 2018, 2019), BERT (Bidirectional Encoder Representations from Transformers; see Devlin et al., 2018) or XLNet (Z. Yang et al., 2019), initial successful experiments on prompt-based learning with the so-called transformer-based language models have been conducted (P. Liu et al., 2021; Schick & Schütze, 2021a). Studies have shown that training and fine-tuning very large language models (e.g. with 10B parameters) is often difficult for independent researchers, requires high computing power and thereby incorporates a high carbon footprint (Schick & Schütze, 2021b). In contrast, using PLMs with prompt-based learning can make predictions without task-specific fine-tuning (X. Liu et al., 2021) and in a low resource setting (T. Gao et al., 2021).

Originating in text generation tasks, prompt-based learning for classification-based tasks has only recently become popular in NLP (P. Liu et al., 2021). One of the first experiments was conducted by Puri and Catanzaro (2019) on GPT-2. They investigated the model's performance

using a zero-shot approach for six different classification tasks. Using prompt-based learning, the performance improved up to 45% compared to random or majority class baselines.

In another study, Yin et al. (2019) combined a so-called ‘BERT model’ (Devlin et al., 2018) for binary text classification with zero-shot prompting. To do so, they trained the model on entailment tasks taken from the datasets ‘MNLI’ (Williams et al., 2018), ‘GLUE RTE’ (A. Wang et al., 2019) and ‘FEVER3’ (converted into dichotomous cases; Thorne et al., 2018). In an entailment task, a model learns a directional relation between text fragments, where the relation holds whenever the truth of one text fragment follows from another text. They then evaluate the model’s performance on three datasets focusing on topic categorization (Yahoo dataset; Zhang et al., 2015), emotion detection (Bostan & Klinger, 2018), and situation frame detection (Strassel et al., 2017). First, they evaluated the model based on partially-unseen labels (i.e. the training is on some labels and the testing on all labels), then experimented on a label-fully-unseen evaluation setting (i.e. no labeled data for task-specific training has been seen before). For this, they used two datasets on ‘topic categorization’ and ‘emotion detection’. Within the label-partially-unseen evaluation, ‘Binary-BERT’ performed remarkably for seen labels, however not for unseen labels. On the third dataset ‘situation frame detection’ being a multi-label classification, the entailment models (pretrained on MNLI) predict better than on the first two datasets. They show a similar performance to ‘Binary-BERT’ for seen labels and an even higher performance regarding unseen labels. Within the label-fully-unseen evaluation, the ‘Wikipedia-based’ baseline model performed much better on the ‘topic categorization’ than on the ‘emotion and situation detection’ tasks. The authors’ explanation is centered on the fact that the Yahoo-based ‘topic categorization’ task is more similar to the Wikipedia-based ‘topic categorization’ problem. However, the authors provide evidence that the pretrained entailment models perform on all three zero-shot text classifications even better, demonstrating these models’ power.

In 2020, Brown et al. investigated the performance of the GPT-3 language model by using few-shot prompts. They reached an accuracy of .864 for a cloze task (LAMBDA dataset; Paperno et al., 2016) or with an accuracy score of .712 for a question-answering task (TriviaQA dataset; Joshi et al., 2017). They showed that this approach results in high performance on the given datasets and tasks such as translation (WMT’14 Fr→En, WMT’16 De→En, and WMT’16 Ro→En), question-answering, and cloze tasks, only leveraging a textual prompt and a few task demonstrations as the input.

Motivated by these promising results, T. Gao et al. (2021) considered classification and regression tasks. By using the small transformer ‘RoBERTa’ and only a few annotated examples (few-shot prompted) they obtained an average improvement of 11% and an absolute improvement

of 30% on provided tasks and outperformed their baselines of standard fine-tuned language models.

Taking into consideration the above-mentioned disadvantages of large language models, Schick and Schütze (Schick & Schütze, 2021a) applied prompt-design on a text classification task and combined it with fine-tuning a set of smaller language models. Thereby, they at first, designed several zero-shot classification prompts, which they applied on each training example, and then fine-tuned several language models, each with the results from one of the zero-shot prompts. The results from these different models are then combined to a single label prediction for each classification (the latter is also known as an ensemble modeling approach; Rokach, 2010). Finally, the predicted labels are then used to train or fine-tune a language model for the actual classification. In the same paper, Schick and Schütze also proposed an iterative version of this process, repeating the steps with increasing sizes of the training data. In a further paper, Schick and Schütze (2021b) then showed that the standard as well as the iterative approach both outperform GPT-3 on the SuperGLUE benchmark task with 32 training examples, although only 0.1% of GPT-3's parameters were used.

Transformer models have been recently used in educational research (Beseiso et al., 2021). The current developments highlighted in the Related Work section underscore the need for further research in education. It is important to find out what impact transformers might have on educational research and practice. However, with regard to prompt-based learning for text classification tasks in educational contexts, which will be the focus of this paper, there have been—to the best of our knowledge—no studies yet.

7.4 Methodology

NLP with transformers in a nutshell

Transformers are machine-learning models originally designed for natural language processing tasks. They can be simplified as probabilistic ‘repetition machines’ that search for patterns learned from tremendous amounts of data and, depending on the task it is trained for, return a numerical representation or a word that corresponds to the learned pattern. In order to do so, they use an attention mechanism, which replaces the recurrent structure of neural networks previously prevalent in machine-learning models (Vaswani et al., 2017). One of its key aspects is the ability to process text data in a parallelizable way, where the relevance of each word (or part of a word) is calibrated in parallel, while recurrent neural networks (RNN) rely on a sequential processing of the text input. In this way, transformers do not suffer from the so-called vanishing gradient problem, which makes it very hard for RNNs to model the relationship of

words that are very far from each other in the text (Hochreiter, 1998; Hochreiter & Schmidhuber, 1997). As a result, transformers enable the learning of much longer text sequences and also allow for a better general representation of natural language. The models are pre-trained in an unsupervised way, where they are trained to predict words (or subwords) that are randomly removed (masked) from the training data. This allows models to be pre-trained with large amounts of text data (Chan et al., 2020). As indicated above, the units (typically called tokens) that the models are trained with are not just words but often also subwords or even just characters (Wu et al., 2016).

The original transformer model (Vaswani et al., 2017) was developed for a translation task and included an encoder part and a decoder part, where the encoder converts an input sequence from the source language into a numerical representation, while the decoder generates a token from the target language based on the numerical input from the encoder and a second input sequence including tokens for the target language that were already generated. Today, the encoder and decoder parts are also used independently. Encoder models such as BERT (Devlin et al., 2018; Sanh et al., 2019) are used for sentence classification or question answering. Decoder models such as the GPT models (Brown et al., 2020; B. Wang & Komatsuzaki, 2021) are suited for text generation. Encoder-Decoder models (similar to the original transformer architecture) are used for summary generations, translation, or generative question answering (Lewis et al., 2019; Y. Liu et al., 2019; Raffel et al., 2019). However, despite architectural differences that make the models particularly suited for certain tasks, most transformer models can be used for a variety of NLP tasks. For example, the architecture and training of GPT models is actually focusing on text generation, by changing the model head they can also be used for text classification or semantic search tasks. And even though the models' architecture and initial training is not focusing on, for example, a classification task, the pure size of the models and the amount of data they were trained with allows them to compete with other smaller models specifically developed for classification tasks (X. Liu et al., 2021). In 2020, researchers from OpenAI (Kaplan et al., 2020) have shown that there is a linear relationship between model performance and model size (including a corresponding increase in training data), and since then there has been an ongoing trend to train larger models. GPT-3 (Brown et al., 2020), was one of the first very large models with 175 billion trained parameters, and in 2021, Microsoft and NVIDIA collaborated to bring out the Megatron-Turing Natural Language Generation model (MT-NLG) with 530 billion parameters (Says, 2021). Since fine-tuning these models is no longer possible without sophisticated computational power, the design of task specific prompts has become a popular alternative. Radford et al. (2018, 2019) and others (Jiang et al., 2020; Petroni et al., 2019)

have shown that prompt-based learning can perform similarly or identically to fine-tuned models on various NLP tasks by providing appropriate natural language prompts.

What is prompt-based learning?

Traditional supervised learning is based on training a model with a large number of input examples and corresponding outputs (labels) for a specific task. Prompt-based learning is based on language models that are trained to predict a word (or subword) for a gap in a natural language text (P. Liu et al., 2021). In this approach, the original input is transformed into a text prompt using a template including predefined gaps, and the language model is then used to probabilistically fill the gaps to obtain a complete text from which the final output can be inferred. In this way, we can manually build a prompt based on natural language to adapt a model to a specific task. When defining a prompt, one may use ‘few-shot’ learning, that is, provide a few examples showing the task (usually not more than a handful; Snell et al., 2017) or use ‘zero-shot’ learning, that is, the model is merely provided with an instruction on the type of expected output data (Romera-Paredes & Torr, 2015). Both approaches allow for the adaptation to new scenarios with little or no labeled data. Prompts follow a *show and tell* character, including instructions, examples, or both. For example, by providing GPT-J with the following text generation prompt¹⁸:

What is Artificial Intelligence in Education? Artificial Intelligence in Education is...

The model will return an answer such as:

Artificial Intelligence in Education is a type of technology that is used in the education system. This is because of the fact that it can be used to solve problems that the traditional systems cannot. However, it is important to understand that this is a complex subject. In this article, we will be talking about what AI is and how it is being used in education. What is Artificial Intelligence?

In this response, the model appears to mimic an article. This makes sense since the model was trained on large amounts of text including many articles. But if we run the prompt again, the response will be different from the previous one and might mimic something else (e.g. a novel,

¹⁸ Try for yourself within this free-of-charge playground by Eleuther.AI: <https://6b.eleuther.ai/> (please use standard settings: TOP-P = .9, Temperature = .8)

a scientific paper, etc.) due to the probabilistic nature of language models. In this section, we will now proceed step-by-step to illustrate a prompt for classification. Besides the formulation of the natural language prompt, it is necessary to control for the degree of determinism that the model has in generating its answer. This is done via a special set of model parameters.

Parameters

The available parameters to control for the model output may differ from model to model but in general all models include at least the following two important parameters: (1) Temperature, which controls the randomness of the predictions, and (2) top-p/top-k, which controls the probability threshold for the prediction. A higher value for temperature means more randomness in the selected prediction (and vice versa), while a smaller value for top-p widens the sampling distribution (i.e. for words that are considered as possible candidates for the prediction) for the prediction (and vice versa). Depending on the task and model in use, it is useful to experiment with these settings to find the best results. Since we aim at a classification task, we decreased the amount of creativity of the LM to a minimum for our study to achieve reliable predictions. Thus, we choose the lowest possible value for temperature and a high value for top-p.

top-p = .9

temperature = 0

Prompt types

As already described above, there are two main types of prompts that can be distinguished: zero-shot prompts and a few-shot prompts. For our experiment, we will make use of both approaches. Furthermore, prompts can be divided into cloze prompts and prefix prompts (P. Liu et al., 2021), where a cloze prompt describes a prompt including a gap in the middle of a text which is to be completed by the model, while a prefix prompt describes one in which the completion follows the given prompt. For classification, we will focus on a prefix prompt. There are several models for zero-shot classification tasks (see *Models - Hugging Face*, n.d.). For the classification of short text answers, a zero-shot approach has been shown to provide satisfactory results in classification tasks, such as determining the sentiment of reviews (Zhong et al., 2021). Therefore, we use a zero-shot classification model as a baseline and will investigate to what extent a few-shot approach will increase the reliability of the prediction.

Prompt engineering

Finally, we have to formulate a template for the prompt. There are several important aspects that need to be considered in the prompt design: the instruction, the example(s), and a blank space for the next word generation. At the beginning of the prompt, we give an instruction for the task: ‘Classify the mail.’ In the given classification task, it is followed by listing the labels that are to be used. In our study, the texts are to be classified according to their politeness, therefore the chosen labels are ‘impolite’, and ‘polite’¹⁹. Then, we separate the instruction from the first example (‘#####’) and provide in total three examples, where each example is clearly separated by the next one (the chosen characters ‘#####’ may also be replaced by other similar characters). Finally, after the last example (and separator), the actual text that is to be classified is provided, followed by ‘label:’ on the next line. Since the model is trained to predict the next probable word in a given text sequence, it will now predict with a very high probability the word ‘polite’ or ‘impolite’ to follow the given pattern and hopefully even choose the correct label to be in accordance with the examples labeled before.

Technically, this is a triple-shot prompt with three examples before the actual classification task. In order for the classification to be correct, the selection of the chosen examples is critical (P. Liu et al., 2021). The examples should not only include correct labels but be as illustrative as possible, for example, they should have a show-and-tell character and contain corner cases. Furthermore, the number and order of examples also affect the output. An exemplary prompt might look like the following:

```
Classify the mail. Labels: impolite, polite
#####
mail: Hello Mrs. Neumann, I recommend an external purchase. With kind regards
label: polite
#####
mail: Good day, unfortunately, I can not do this task either !!!!!!!!!!!!!!!!!!!!!!!
label: impolite
#####
mail: Hello Mr. Brown, can you not even solve a task alone?
label: impolite
#####
mail: Dear Mrs. Miller, I do not like the way this works. Do it on your own.
```

¹⁹ An alternative instruction could, for example, also be ‘Classify the text using the labels impolite and polite.’ In general, it is not possible to tell which exact formulation will work best since this also depends very much on the underlying language model and the texts it was trained on.

label:

With the correct parameter setup this should result in the prediction of ‘impolite’ as the next word. Depending on the model we are using and the task we are performing, other prompts may be more appropriate due to differences in training data, model sizes, etc. In fact, research is currently being done on automating prompts to make the engineering process of prompts easier and more dynamic (Shin et al., 2020). While the last mail could be manually exchanged with all mails that are to be classified, this will usually be done via a script, looping over all given mail texts, replacing the mail classified before executing the prediction.

7.5 Experiment

Dataset

The experiment for the presented study is based on a dataset of originally 2,088 short email responses (averaging about 62 words) from 780 trainees assessed with an authentic office simulation. Seventy percent ($n = 1461$) of the responses were used as training data for the fine-tuning approach (see also Ludwig et al., 2021); 30% ($n = 627$) were used as test data for the fine-tuning as well as for the prompt-based approach. All trainees were in the middle or final phase of a three-year commercial apprenticeship. Participants were asked to complete three complex domain-specific tasks from the field of controlling and to communicate their decision appropriately by mail. Part of the domain-specific competence model was communication style, which was manually labeled (among other items) as polite or impolite. The underlying competency model addresses appropriate communication as follows: “A competent problem-solver communicates his or her solution in an adequate manner to external or internal stakeholders” (Rausch & Wuttke, 2016, p. 177). The majority of all mails were rated as polite by several coders (see Brandt et al., 2016; Rausch et al., 2016; Seifried et al., 2020; Sembill et al., 2016). We split the data into a test (30%) and training dataset (70%) for further processing.

Models

As already mentioned above, we use zero-shot classification as a baseline comparison. The application of zero-shot classifications was examined by Yin et al. (2019) as a promising method for using pre-trained models. Zero-shot models can yield surprisingly high levels of accuracy in classifying tasks, for example, to determine the general topic of a text (Pushp & Srivastava, 2017). For the zero-shot approach, we use the pre-trained model bart-large-mnli, a BART model (Lewis et al., 2019) trained by meta (formerly Facebook AI). BART is a trans-

former model with a bidirectional (BERT-like) encoder and an autoregressive (GPT-like) decoder part containing 406 million parameters trained on 160 GB of data, including news, books, stories, and web texts (Lewis et al., 2019; Y. Liu et al., 2019). For the actual study, we use GPT-J-6B, which runs on a self-deployed instance in Colab. The model can be used for few-shot classification and have been trained on a large amount of data. GPT-J-6B has more than six billion parameters trained on ‘The Pile’, an 825 GB dataset merged from 22 datasets, including academic and professional sources such as PubMed, FreeLaw, Wikipedia, and Github (L. Gao et al., 2020). In fact, GPT-J-6B was developed as an open-source alternative to GPT-3.

Table 7-1 provides an overview of the three used models. Prompt examples can be found in the Appendix. We refer all interested readers to our public repository, including notebooks with further instructions on the usage of the models, code, comments and model specific variations (see reference to the GitHub Repo in the Appendix).

Table 7-1: Overview of the models for the experiment.

Model	Parameters	Training data	Prompt type	Free-of-charge
BART-large-mnli	406M	160Gb	zero-shot	Yes
GPT-J-6B	6B	825Gb	few-shot	Yes

Recoding

In order to gain more insights into the reliability and potential problems in the human ratings (research question 3), the responses of the test data set were rated again by two independent human raters. The coding manual used was based on that of the original study but was expanded to include more thorough classification examples. The labels were then further improved by resolving disagreements via discussions between two raters. When in doubt, the particular mail was considered polite in the student's favor. The additional examples for coding guide were the following:

- (1) Rude mails that contain offensive language, for example mail #93: “Please find another douche bag to do your dirty work! With kind regards”.
- (2) Facetious mails that make use of inappropriate humor, for example mail #305: “Attached you will find the current Wikipedia article about the Incoterms, with which you

certainly can do nothing. I would be happy to send you more useless links, which, like the attached file, have no informational content whatsoever”.

- (3) Excuse mails that shift the blame to other circumstances, for example mail #118: “Good day Mrs. Meier, I am not satisfied with the functions of your Excel application, if you improve them you can contact me again”.²⁰

7.6 Results

To compare the different models and prompt types and provide an answer to research question 1, we calculate the following performance metrics: Accuracy, F1 score, ROC AUC (Zou et al., 2007), and Cohen's Kappa (Cohen, 1968). The results on the performance metrics and the corresponding confusion matrices are shown in Table 7-2 and Table 7-3, respectively. The revised test dataset (and no training dataset) is used for prompt-based learning (see RQ3 results). The results for the zero-shot baseline model are the lowest across all metrics ($Acc = .80$, $F1 = .89$, $AUC = .64$, $k = .16$) with, in total, 126 incorrect predictions (about 20%), mainly for false predicted impoliteness.

The GPT-J-6B model with a few-shot prompt providing four examples (see Appendix for the final prompt) yields significantly better results ($Acc = .93$, $F1 = .96$, $AUC = .66$, $k = .40$) with, in total, 44 incorrect predictions (about 7%).

Table 7-2: Performance per model.

Model	Accuracy	F1-Score	ROC AUC Score	Kappa
BART-large-mnli	.80	.89	.64	.16
GPT-J-6B	.93	.96	.66	.40

²⁰ Mails were translated from German into English for better reading flow.

Table 7-3: Confusion matrix per model.

Actual / Predicted	Impolite	Polite
<i>BART-large-mnli</i>		
Impolite	22	28
Polite	98	479
<i>GPT-J-6B</i>		
Impolite	17	33
Polite	11	566

To illustrate the performance differences between the prompt design approach and the fine-tuning approach (research question 2), we used the German BERT base model (Chan et al., 2020), a BERT model pre-trained particularly with German text. We then fine-tuned it with the 70% split of the data (1461 mails) that was not part of the test data used in the analyses for research question 1 above. The test data for the fine-tuned model is therefore equivalent to the one used in the prompt-based approach. Similarly, the same metrics and a confusion matrix are calculated. The performance metrics (Table 7-4) and confusion matrix (Table 7-5) show results similar to the few-shot prompt-design considering the accuracy and the F1 score. The ROC AUC Score and the Kappa values are substantially higher though for the fine-tuned model.

Table 7-4: Performance metrics for fine-tuning.

Model	Accuracy	F1-Score	ROC AUC Score	Kappa
German BERT (Oct. 2020)	.92	.96	.88	.59

Table 7-5: Confusion matrix for fine-tuning.

Actual / Predicted	Impolite	Polite
<i>German BERT base (Oct. 2020)</i>		
Impolite	41	9
Polite	39	538

To gain more insights on how machine ratings can be used to improve human ratings, we first further investigated the reliability of the human ratings. We found that the original human ratings and the newly conducted ratings resulted in a Cohen’s Kappa of .58.

Furthermore, machine ratings have the option of additionally providing a reliability estimate for each single rating, indicating how sure the model is that its estimated classification is correct. Ludwig and colleagues (2021) have shown that the fine-tuned BERT model classified 8 ratings with a probability of 95% or higher in a class that was different from the rating of the human raters. A revision of the responses showed that original human ratings were incorrect in these cases²¹. In a similar way, the results of the prompt-based model predictions can be investigated considering the reliability of their individual classifications. Unfortunately, the used APIs for the prompt-based approaches currently do not provide sufficient detailed access to such results (however, they are expected to do so in the very near future). The inter-rater reliability between original human ratings and a second human rating ($k = .58$) can be considered moderate (Landis & Koch, 1977).

7.7 Discussion

The application of zero-shot and few-shot prompt-based learning with transformer-based language models is a still novel method in the field of education. We evaluate machine ratings alongside a human rater and apply the method to a real-world classification task, namely identifying whether a commercial trainee wrote a polite or impolite business mail.

To answer research question 1, to what extent a learning approach based on zero and few-shot prompts with current language models (bart-large-mnli, GPT-J-6B) can provide reliable results

²¹ Unfortunately, for privacy reasons it is not possible to share them.

for a classification task, we use a baseline model and a current state-of-the-art language model. We find that the baseline model, a learning approach based on zero-shot prompts, provides unsatisfactory results compared to human ratings. This may be related to the specific background that might not be considered when training the model. In this classification task, longer texts with a domain-specific background are processed. Previous Research on Zero-Shot Classification (Pushp & Srivastava, 2017; Sainz & Rigau, 2021; Yin et al., 2019; Zhong et al., 2021) were often performed at the sentence level or on very short texts (often tweets). Despite these shortcomings, with 80% correctly rated responses, the baseline model is still able to correctly distinguish between polite and impolite for the majority of the mails. However, the overall low agreement between human and machine rating is unsatisfactory ($k = .16$).

We then continued the experiment with four to eight examples in a few-shot prompt learning approach, which significantly increased the agreement with the human ratings, reaching a Cohen's Kappa of .40 using the GPT-J model. In General, this will still be considered too low though to completely rely on such prompt-based ratings. Further, the achievable agreement is not easily transferable to other tasks, which might be harder or easier to predict by providing a selected set of examples. Therefore, this approach will always need to be applied alongside human raters to ensure high reliability.

To answer research question 2, to what extent does prompt-based learning provides comparable results to a fine-tuned language model, we compare the best few-shot prompt-based learning approach (which provides few examples in one prompt to GPT-J) with a German-language BERT model fine-tuned with a training dataset of 1461 labeled mails. The results show that the approach performs better ($k = .59$ for BERT) than prompt-based learning with a few-shot approach ($k = .40$). However, a Cohen's Kappa of .59 will probably still be considered too low to completely rely on such machine ratings. This is in line with previous findings, for example, Yang et al. (2017) reported a kappa of .625 for short text grading with support vector machine (SVM) and long short-term memory (LSTM). Similar, Ndukwe et al. (2019) reported a kappa of .6 for a machine learning grading using chatbots. That is, just as in the case of the prompt-based approach, it will still be necessary to apply the approach alongside human raters. Further, while the fine-tuning approach is able to classify about 12% more responses correctly than the few-shot prompt-based approach, this increase comes at the cost of having to manually rate a training set, in this case with 1.461 responses, which greatly increases the burden of necessary human ratings.

Finally, considering research question 3, to investigate the potential of machine ratings to improve human ratings, we found that a new human rating based on a corresponding scoring

guideline resulted in a Cohen's Kappa of .58 for the agreement between the old and the new human ratings, which provides an indication of the subjectivity of human ratings, in particular when the variety of response cases is too large to provide sufficient examples to the raters to cover the full variety. Previous research shows that human ratings can vary depending on task and text length, for example, Lind Pantzare (2015) reports lower kappa values for human ratings in extended answers than for short texts or multiple-choice questions. Human-to-human correlations sometimes may even worse than human-machine correlations (Ramineni et al., 2012; see also Shermis & Burstein, 2013). In all three considered approaches the vast majority of the responses seem to be easy to rate correctly. Quality control for large scale assessments shows that an agreement between raters of over 70% can often still be considered as acceptable (Becker, 2020). However, 10 to 20% of the responses seem to be more difficult to rate, and here it is where the machine ratings as well as the human ratings differ from each other.

The considered experiment still has some limitations: The reported results are only for a single rating task, the agreement of the machine and human ratings will typically depend very much on the complexity and subjectivity of the given task. In this experiment, we considered a rather subjective task by classifying mails into polite or impolite. We would like to emphasize that the guidelines established by Landis and Koch (Landis & Koch, 1977) can do more harm than good. Landis and Koch provide no empirical evidence for their thresholds. In addition, we also calculated Krippendorff's alpha ($\alpha = .71$) which can be considered acceptable (Krippendorff, 2004). Originally, we also ran tests with paid language models, but the results were only slightly better, so we stuck with the free models to promote the democratization of AI and the use of free language models, and to encourage other researchers to work with public available transformer models.

7.8 Conclusion - Reciprocal Perspectives on AI and Human Intelligence in Education

Based on these results, we would like to highlight and discuss the use of transformer-based language models as an alternative second machine evaluation in a reciprocal work loop with humans. Prompt-based learning with language models can be a powerful approach to classification tasks that complements human raters. This methodological approach has advantages, especially compared to fine-tuning language models and various other machine learning models, but it also poses some risks. A typical training dataset for fine-tuning contains several thousand data points. Except for very large assessments, the amount of training data necessary might therefore easily be as large as the full response data set, or, in fact, maybe even larger than the

collected response data. Prompt-based learning can provide almost comparable results without large amounts of training data. By using those response ratings from a prompt-based approach, for which the model yields at least a 95 % probability for the predicted response category, the prompt-based approach allows to rate the vast majority of items automatically and allows the human raters to focus exclusively on the remaining responses, that are difficult to rate (i.e. where the probability of the predicted class is lower than 95 %). A major advantage of using the prompt-based approach is then the overall lower cost of data labeling, first, because no further training of a model with labeled training data is required, and second, due to the smaller amount of response that need to be labeled by humans.

A further significant advantage can be that due to the much smaller amount of responses that need to be rated by humans, each of these responses can be rated or maybe even discussed by several human raters to get overall to significantly more reliable ratings. The approach should always be used judiciously as a ‘human in the loop’ approach, but it has the potential to reliably predict categorization for fairly clear cases and to focus the attention of humans for the remaining unclear and ambiguous corner cases.

Nevertheless, there are also limitations and drawbacks for the prompt-based approach: If sufficient training data is available, the fine-tuning approach will probably still allow for labeling more responses automatically correct than the prompt-based approach. Developing prompts is still a challenging task that takes a lot of time and should not be underestimated. The development of an appropriate prompt can be viewed as an exploratory and iterative process. This process should be done judiciously with the involvement of a second human evaluator to review the results of the model. This is similar to standard research practice. When two independent human raters are faced with a similar task, they often need to be guided by a detailed coding manual, which may be revised after an initial round of ratings. Some of the latest major language models are only accessible through APIs and charge fees for their services, which leads to additional costs. Finally, although the models we use are mostly multilingual, English is still the majority class when training language models. Therefore, we expect the results to be much better in English. More importantly, we emphasize the need for more democratization of AI to train more multilingual models, not only for German, but also for minority languages.

Besides the technicalities of achieving reliable response ratings, it is also important to consider and highlight the importance of fairness and data privacy, when applying artificial intelligence in an educational setting. First, handling sensitive educational data requires privacy protection. Researchers must meet privacy-related requirements because student privacy must be respected, especially when inferences about achievement or behavior are made. In this work, no

inferences relevant to students were made and responses were pseudonymized. In addition, we would like to emphasize that the experiment did not involve a graded performance test or similar educational assessment. However, if inferences are made about performance, it is essential to inform students in advance about the use of algorithms to ensure their acceptance and also to allow for rejection. Previous research has shown that full transparency is required for acceptance, whether algorithms are used for simple processing or are implemented in learning analytics systems (Ifenthaler & Schumacher, 2016; Ifenthaler & Tracey, 2016). Confident use of data includes concern about bias in AI algorithms (Pethig & Kroenung, 2022). Despite the advantages of transformers, it is still difficult to work with unbiased algorithms (Silva et al., 2021). Several publications show that, for example, GPT-3 tends to discriminate against religious groups (Abid et al., 2021) or gender (Lucy & Bamman, 2021). Furthermore, we would like to emphasize that the actual use in educational practice remains unclear and an ethical debate is urgently needed. In light of the findings of this study, the current state of transformational models is far from being applied in educational research and practice, and further research is needed, especially in the area of text classification using prompt-based learning. For example, one could debate whether a model such as the one proposed should be used as an assistance system for students (by displaying a warning on the screen before sending an email about possible rudeness) or for teachers as an assistance system. It could be very difficult to make real decisions and assign grades without the involvement of teachers and the consent of stakeholders.

Second, we would like to highlight the practical impact that transformers can have. Since prompt-based learning is still in its infancy, we believe that this approach has a high development potential that will be further uncovered in the coming years. Work is already underway in this area, for example, on automatic prompting (Shin et al., 2020), on more technical methods of matching prompts without discrete prompts (Lester et al., 2021), and the combination of prompt-based learning and fine-tuning of large and even smaller language models (Schick & Schütze, 2021a, 2021b). There could be different looking prompts that serve the same purpose or work even better - we just don't know. Prompt engineering is, at least for now, more of an art than a science, and more research is needed to explore prompt engineering. To solve this problem, repositories such as PromptSource have recently been introduced to create, share and use prompts more systematically. The reciprocal two-way workflow between humans and artificial intelligence is a matter of give-and-take which we would like to underline. Artificial intelligence can be updated through prompts. This is particularly interesting for new and previously unknown scenarios (e.g. when a model needs to learn about the 'coronavirus') and for domains that are more specific and less common. Scao and Rush show that a single prompt can

be worth 100 data points at a time (2021). Human intelligence is required to instruct the language models with a particular task. Without human instruction, these models are fairly useless, despite the large corpora on which they are trained. Although there are also several technical modifications to improve prompts, we suggest experimenting for a while with a smaller amount of data to find the best natural language prompt for a model. Depending on the task and model, for example, initial instruction can lead to better or worse results. However, the potential of language models can be brought to a wider audience through the use of prompts. Prompt-based learning is a low-code and potentially even no-code approach that lowers the barrier for educational researchers to be able to implement AI in their research. We therefore think it has the potential to become a standard practice for applying AI in educational research as well as in educational practice. A continuing trend to train large transformational models undermines this potential. Recently, Microsoft and NVIDIA have collaborated to work on a 530 billion parameter model and recently published the Megatron-Turing Natural Language Generation (MT-NLG) model and models like the recently announced GPT-4 are expected to be bigger than ever (with over 100 trillion parameters). We therefore expect that these models will be even better in understanding prompt-based instructions and providing reliable response ratings. Further, grassroots research collectives like Eleuther.AI and BigScience are trying to keep up and preserve AI models as open source (Leahy, 2022), which gives us hope that language models will also be available at an affordable price in the future.

7.9 References

- Abid, A., Farooqi, M., & Zou, J. (2021). Persistent anti-muslim bias in large language models. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 298–306.
- Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V. 2. *The Journal of Technology, Learning and Assessment*, 4(3).
- Becker, A. (2020). Post-collection Data Capture, Scoring, and Processing. In H. Wagemaker (Ed.), *Reliability and Validity of International Large-Scale Assessment* (pp. 151–167). Springer. https://doi.org/10.1007/978-3-030-53081-5_9
- Beseiso, M., Alzubi, O. A., & Rashaideh, H. (2021). A novel automated essay scoring approach for reliable higher educational assessments. *Journal of Computing in Higher Education*, 33(3), 727–746. <https://doi.org/10.1007/s12528-021-09283-1>
- Bin, L., & Jian-Min, Y. (2011). Automated Essay Scoring Using Multi-classifier Fusion. In Y. Wu (Ed.), *Computing and Intelligent Systems* (pp. 151–157). Springer. https://doi.org/10.1007/978-3-642-24010-2_21
- Bostan, L.-A.-M., & Klinger, R. (2018). An Analysis of Annotated Corpora for Emotion Classification in Text. *Proceedings of the 27th International Conference on Computational Linguistics*, 2104–2119. <https://aclanthology.org/C18-1179>
- Brandt, S., Rausch, A., & Kögler, K. (2016). *A Scoring Procedure for Complex Assessments Focusing on Validity and Appropriate Reliability*. 1–20.

- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., & Askell, A. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Chan, B., Schweter, S., & Möller, T. (2020). German’s Next Language Model. *Proceedings of the 28th International Conference on Computational Linguistics*, 6788–6796.
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. <https://doi.org/10.1037/h0026256>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. <https://doi.org/10.48550/ARXIV.1810.04805>
- Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., & Nabeshima, N. (2020). The pile: An 800gb dataset of diverse text for language modeling. *arXiv Preprint arXiv:2101.00027*.
- Gao, T., Fisch, A., & Chen, D. (2021). Making Pre-trained Language Models Better Few-shot Learners. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 3816–3830.
- Gillioz, A., Casas, J., Mugellini, E., & Khaled, O. A. (2020). Overview of the Transformer-based Models for NLP Tasks. *15th Conference on Computer Science and Information Systems (FedCSIS)*, 179–183. <https://doi.org/10.15439/2020F20>
- Guan, C., Mou, J., & Jiang, Z. (2020). Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. *International Journal of Innovation Studies*, 4(4), 134–147. <https://doi.org/10.1016/j.ijis.2020.09.001>
- Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., & Xu, Y. (2022). A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(1), 87–110.
- Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(2), 107–116.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Ifenthaler, D. (2014). AKOVIA: Automated Knowledge Visualization and Assessment. *Technology, Knowledge and Learning*, 19, 241–248. <https://doi.org/10.1007/s10758-014-9224-6>
- Ifenthaler, D. (2022). Automated Essay Scoring Systems. In O. Zawacki-Richter & I. Jung (Eds.), *Handbook of Open, Distance and Digital Education* (pp. 1–15). Springer. https://doi.org/10.1007/978-981-19-0351-9_59-1
- Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development*, 64(5), 923–938. <https://doi.org/10.1007/s11423-016-9477-y>
- Ifenthaler, D., & Tracey, M. W. (2016). Exploring the relationship of ethics and privacy in learning analytics and design: Implications for the field of educational technology. *Educational Technology Research and Development*, 64(5), 877–880. <https://doi.org/10.1007/s11423-016-9480-3>

- Jiang, Z., Xu, F. F., Araki, J., & Neubig, G. (2020). How Can We Know What Language Models Know? *Transactions of the Association for Computational Linguistics*, 8, 423–438.
- Joshi, M., Choi, E., Weld, D. S., & Zettlemoyer, L. (2017). TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension. *arXiv:1705.03551 [Cs]*. <http://arxiv.org/abs/1705.03551>
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling Laws for Neural Language Models. *arXiv:2001.08361 [Cs, Stat]*. <http://arxiv.org/abs/2001.08361>
- Ke, Z., & Ng, V. (2019). Automated Essay Scoring: A Survey of the State of the Art. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 6300–6308. <https://doi.org/10.24963/ijcai.2019/879>
- Krippendorff, K. (2004). Reliability in Content Analysis.: Some Common Misconceptions and Recommendations. *Human Communication Research*, 30(3), 411–433. <https://doi.org/10.1111/j.1468-2958.2004.tb00738.x>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Leahy, C. (2022, February 2). *Announcing GPT-NeoX-20B*. EleutherAI Blog. <https://blog.eleuther.ai/announcing-20b/>
- Lester, B., Al-Rfou, R., & Constant, N. (2021). The Power of Scale for Parameter-Efficient Prompt Tuning. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 3045–3059.
- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv Preprint arXiv:1910.13461*.
- Liang, G., On, B.-W., Jeong, D., Kim, H.-C., & Choi, G. (2018). Automated Essay Scoring: A Siamese Bidirectional LSTM Neural Network Architecture. *Symmetry*, 10(12), 682. <https://doi.org/10.3390/sym10120682>
- Lind Pantzare, A. (2015). Interrater reliability in large-scale assessments—Can teachers score national tests reliably without external controls? *Practical Assessment, Research, and Evaluation*, 20(1), 9.
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2021). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *arXiv:2107.13586 [Cs]*. <http://arxiv.org/abs/2107.13586>
- Liu, X., Ji, K., Fu, Y., Du, Z., Yang, Z., & Tang, J. (2021). P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning Universally Across Scales and Tasks. *arXiv Preprint arXiv:2110.07602*.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv:1907.11692 [Cs]*. <http://arxiv.org/abs/1907.11692>
- Lucy, L., & Bamman, D. (2021). Gender and Representation Bias in GPT-3 Generated Stories. *Proceedings of the Third Workshop on Narrative Understanding*, 48–55. <https://doi.org/10.18653/v1/2021.nuse-1.5>
- Ludwig, S., Mayer, C., Hansen, C., Eilers, K., & Brandt, S. (2021). Automated essay scoring using transformer models. *Psych*, 3(4), 897–915.
- Models—Hugging Face*. (n.d.). Retrieved March 30, 2022, from <https://huggingface.co/models>

- Ndukwe, I. G., Daniel, B. K., & Amadi, C. E. (2019). A Machine Learning Grading System Using Chatbots. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Artificial Intelligence in Education* (pp. 365–368). Springer. https://doi.org/10.1007/978-3-030-23207-8_67
- Paperno, D., Kruszewski, G., Lazaridou, A., Pham, N.-Q., Bernardi, R., Pezzelle, S., Baroni, M., Boleda, G., & Fernández, R. (2016). The LAMBADA dataset: Word prediction requiring a broad discourse context. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1525–1534.
- Papers with Code—The latest in Machine Learning*. (n.d.). Retrieved March 21, 2022, from <https://paperswithcode.com/>
- Pethig, F., & Kroenung, J. (2022). Biased Humans, (Un)Biased Algorithms? *Journal of Business Ethics*, 183, 637–652. <https://doi.org/10.1007/s10551-022-05071-8>
- Petroni, F., Rocktäschel, T., Riedel, S., Lewis, P., Bakhtin, A., Wu, Y., & Miller, A. (2019). Language Models as Knowledge Bases? *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2463–2473.
- Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*, 12, 22. <https://doi.org/10.1186/s41039-017-0062-8>
- Puri, R., & Catanzaro, B. (2019). Zero-shot Text Classification With Generative Language Models. *arXiv:1912.10165 [Cs]*. <http://arxiv.org/abs/1912.10165>
- Pushp, P. K., & Srivastava, M. M. (2017). Train once, test anywhere: Zero-shot learning for text classification. *arXiv Preprint arXiv:1712.05972*.
- Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained models for natural language processing: A survey. *Science China Technological Sciences*, 63(10), 1872–1897. <https://doi.org/10.1007/s11431-020-1647-3>
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. *OpenAI Blog*.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*.
- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2019). *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*. <https://doi.org/10.48550/ARXIV.1910.10683>
- Ramineni, C., Trapani, C. S., Williamson, D. M., Davey, T., & Bridgeman, B. (2012). Evaluation of the e-rate® scoring engine for the TOEFL® independent and integrated prompts. *ETS Research Report Series*, 2012(1), i–51. <https://doi.org/10.1002/j.2333-8504.2012.tb02288.x>
- Rausch, A., Seifried, J., Wuttke, E., Kögler, K., & Brandt, S. (2016). Reliability and validity of a computer-based assessment of cognitive and non-cognitive facets of problem-solving competence in the business domain. *Empirical Research in Vocational Education and Training*, 8(1), 9. <https://doi.org/10.1186/s40461-016-0035-y>
- Rausch, A., & Wuttke, E. (2016). Development of a Multi-Faceted Model of Domain-Specific Problem-Solving Competence and its Acceptance by Different Stakeholders in the Business Domain. *Unterrichtswissenschaft*, 44(2), 169–184.
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33, 1–39. <https://doi.org/10.1007/s10462-009-9124-7>

- Romera-Paredes, B., & Torr, P. (2015). An embarrassingly simple approach to zero-shot learning. *Proceedings of the 32nd International Conference on Machine Learning*, 2152–2161. <https://proceedings.mlr.press/v37/romera-paredes15.html>
- Sainz, O., & Rigau, G. (2021). Ask2Transformers: Zero-Shot Domain labelling with Pretrained Language Models. *Proceedings of the 11th Global Wordnet Conference*, 44–52.
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv Preprint arXiv:1910.01108*.
- Says, U. (2021, October 11). *Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, the World's Largest and Most Powerful Generative Language Model*. NVIDIA Technical Blog. <https://developer.nvidia.com/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most-powerful-generative-language-model/>
- Scao, T. L., & Rush, A. (2021). How many data points is a prompt worth? *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2627–2636. <https://doi.org/10.18653/v1/2021.naacl-main.208>
- Schick, T., & Schütze, H. (2021a). Exploiting Cloze-Questions for Few-Shot Text Classification and Natural Language Inference. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 255–269.
- Schick, T., & Schütze, H. (2021b). It's Not Just Size That Matters: Small Language Models Are Also Few-Shot Learners. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2339–2352.
- Seifried, J., Brandt, S., Kögler, K., & Rausch, A. (2020). The computer-based assessment of domain-specific problem-solving competence—A three-step scoring procedure. *Cogent Education*, 7(1), 1719571. <https://doi.org/10.1080/2331186X.2020.1719571>
- Sembill, D., Rausch, A., Wuttke, E., Seifried, J., Wolf, K. D., Martens, T., & Brandt, S. (2016). *Domain specific Problem-solving Skills of Prospective Industrial Clerks (DomPL-IK) Modellierung und Messung domänenspezifischer Problemlösekompetenz bei Industriekaufleuten (DomPL-IK)* (Version 1) [Dataset]. IQB - Institute for Educational Quality Improvement. https://doi.org/10.5159/IQB_DOMPL-IK_V1
- Shermis, M. D., & Burstein, J. (Eds.). (2013). *Handbook on automated essay evaluation: Current applications and new directions*. Routledge.
- Shin, T., Razeghi, Y., Logan IV, R. L., Wallace, E., & Singh, S. (2020). AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 4222–4235.
- Silva, A., Tambwekar, P., & Gombolay, M. (2021). Towards a Comprehensive Understanding and Accurate Evaluation of Societal Biases in Pre-Trained Transformers. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2383–2389. <https://doi.org/10.18653/v1/2021.naacl-main.189>
- Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4080–4090.

- Strassel, S. M., Bies, A., & Tracey, J. (2017). Situational Awareness for Low Resource Languages: The LORELEI Situation Frame Annotation Task. *SMERP@ECIR*, 32–41.
- Taghipour, K., & Ng, H. T. (2016). A neural approach to automated essay scoring. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1882–1891.
- Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). *FEVER: A large-scale dataset for Fact Extraction and VERification* (arXiv:1803.05355). arXiv. <http://arxiv.org/abs/1803.05355>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 5998–6008.
- Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2019). GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. *arXiv:1804.07461 [Cs]*. <http://arxiv.org/abs/1804.07461>
- Wang, B., & Komatsuzaki, A. (2021, May). *GPT-J-6B: A 6 Billion Parameter Autoregressive Language Model*. <https://github.com/kingoflolz/mesh-transformer-jax>
- Williams, A., Nangia, N., & Bowman, S. R. (2018). *A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference* (arXiv:1704.05426). arXiv. <http://arxiv.org/abs/1704.05426>
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., ... Dean, J. (2016). Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *arXiv:1609.08144 [Cs]*. <http://arxiv.org/abs/1609.08144>
- Yamamoto, M., Umemura, N., & Kawano, H. (2018). Automated Essay Scoring System Based on Rubric. In R. Lee (Ed.), *Applied computing & information technology* (pp. 177–190). Springer. https://doi.org/10.1007/978-3-319-64051-8_11
- Yang, X., Zhang, L., & Yu, S. (2017). Can Short Answers to Open Response Questions Be Auto-Graded Without a Grading Rubric? In E. André, R. Baker, X. Hu, Ma. M. T. Rodrigo, & B. du Boulay (Eds.), *Artificial Intelligence in Education (AIED)* (pp. 594–597). Springer. https://doi.org/10.1007/978-3-319-61425-0_72
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, 32.
- Yin, W., Hay, J., & Roth, D. (2019). Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3912–3921. <https://doi.org/10.18653/v1/D19-1404>
- Zhang, X., Zhao, J., & LeCun, Y. (2015). *Character-level Convolutional Networks for Text Classification* (Version 3). arXiv. <https://doi.org/10.48550/ARXIV.1509.01626>
- Zhong, R., Lee, K., Zhang, Z., & Klein, D. (2021). Adapting language models for zero-shot learning by meta-tuning on dataset and prompt collections. *arXiv Preprint arXiv:2104.04670*.

Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. *Circulation*, *115*(5), 654–657. <https://doi.org/10.1161/CIRCULATIONAHA.105.594929>

8 Discussion and Research Outlook

In this concluding chapter, the research questions of this thesis are addressed by summarizing the key findings from the five underpinning studies (section 8.1). Following this, the limitations of the research contributions are identified and discussed (section 8.2). The thesis concludes with implications for applying the findings (section 8.3).

8.1 Summary

As stated at the beginning of this dissertation, digitalization continuously drives new opportunities and challenges for economic and business education in VET. It impacts teaching, learning, and VET research (Euler & Wilbers, 2020; Wuttke et al., 2020). To address these challenges, this dissertation *focuses on the professional development of (prospective) teachers and examines how technology can improve educational processes*. The results of the five studies conducted for this dissertation help to answer this overarching question by reflecting three main areas of research. The first research area focuses on the current use of digital tools by teachers in commercial vocational schools. It raises the question of the current state of digital tools in these institutions (Study 1; Mayer, Gentner, et al., 2023) The study examines *how digital tools are currently used in vocational schools and what teachers' expectations are for the further integration of digital tools alongside face-to-face teaching after the pandemic*. The second line of research relates to teaching and learning with digital tools in vocational education and training and explores the question of *how (prospective) teachers can be explicitly trained and supported through the use of sophisticated digital tools*. This research was conducted by implementing a teacher training course on teaching and learning with ERP systems (Study 2; Mayer & Seifried, 2024) and evaluating the course and students ERP skills in a pre-post design, as well as through an experimental application of AI prompting to determine whether a large language model can automatically and reliably score student responses to assist educators (Study 5; Mayer et al., 2022). The third line of research investigates individual learning processes and examines *how the use of unobtrusive measuring instruments, such as eye-tracking, can provide insights into learning processes in digital learning environments*. To this end, the current state of eye-tracking research in computer-based simulations in VET was analyzed in a scoping review (Study 3; Mayer, Rausch, et al., 2023), followed by a first eye-tracking experiment in a computer-based learning environment for economic education (Study 4; Mayer et al., under review).

The current state of digital tools in commercial schools

Study 1 examined how digital teaching practice was carried out in commercial vocational schools during the COVID-19 pandemic. The results show that teachers generally reported positive conditions in terms of technical equipment and organizational support during the pandemic. Most teachers confirmed the presence of sufficient Internet connection and support in acquiring the necessary digital skills. A high percentage of lessons were conducted synchronously. Commonly used digital tools included learning platforms, videoconferencing tools, and cloud solutions for file sharing. Less commonly used tools included instructional videos, presentations, classroom discussions, surveys, quizzes. The study also examines how digital teaching practices in vocational schools can be characterized using the SAMR model. The results show that most of the digital tools used by teachers fell into the substitution and augmentation levels of the SAMR model. These levels indicate that digital tools were primarily used to enhance existing teaching practices rather than to change them. Tools that could lead to a significant change in digital teaching practices, such as those that enable complex teaching and learning arrangements like computer-based simulations or learning analytics, were rarely used. Finally, the study examines teachers' expectations of teaching practice after the COVID-19 pandemic. Analysis of open-ended responses revealed that teachers expect digital tools to continue to play a role in supplementing face-to-face teaching. They expect to continue to use digital tools after the pandemic to provide materials and maintain digital communication channels. This expectation underscores the perceived value and lasting impact of digital tools in enhancing the educational experience.

Teaching and learning with sophisticated digital tools in VET

In a next step, the thesis investigated how to prepare prospective teachers to teach with digital tools, in particular to teach business processes knowledge with ERP systems in vocational schools (Study 2). To this end, a master's course was developed that focused on teaching and learning with ERP systems. The students were very satisfied with the course and reported positive effects on their self-assessed ERP competencies, even though the TPACK dimensions changed only slightly. Reflection journals indicated that some students had overestimated their skills at the beginning of the semester, which could partly explain the small significant changes. Students with lower self-assessments at the beginning of the semester showed significant improvements in technological and subject-specific knowledge, while those with higher self-assessments showed no significant changes. The study also highlighted that difficult problems often arise when working on business tasks within ERP systems due to insufficient basic technical and content knowledge, such as a lack of understanding of ERP system structures and the

application of authentic business processes. In summary, specific training on ERP systems is essential for (prospective) teachers to integrate this technology into the classroom and overcome typical challenges and minimize errors through experience. This training is therefore relevant to the promotion of teaching and learning with ERP systems and a crucial part of designing effective teaching-learning environments with ERP systems in VET.

Another advanced digital tool is AI, which is currently often considered a ‘game changer’ in education. Therefore, this feasibility study focuses on first determining its effectiveness with prompt-based learning in the context of a simple dichotomous classification task (Study 5). The study examined the performance of two state-of-the-art open-source models in classifying the politeness of participants' short text responses to a complex problem-solving task. It specifically assessed their effectiveness using zero- and few-shot prompts. The zero-shot model achieved only 80% accuracy and a low Cohen's Kappa, likely due to the language and complexity of the text. A few-shot approach with four to eight examples improved agreement with human ratings, but reliability was still too low to rely on prompt-based ratings alone. When the prompt-based learning approach was compared with a fine-tuned German-language BERT model trained on emails labeled by humans, the fine-tuned BERT model performed better, achieving a Cohen's Kappa of .59 versus .40 for the few-shot prompt approach. However, fine-tuning requires extensive manual labeling, which increases the workload for humans. The study suggests that AI has the potential to improve human assessment by providing consistent baseline values. A recent human score based on the scoring guidelines yielded a Cohen's Kappa of .58, illustrating the subjectivity of human scoring. AI scoring helps human raters focus on ambiguous responses, improving overall scoring reliability and reducing workload.

Insight into learning processes in digital learning environments

In order to gain insight into learning processes, an overview of the current state of eye-tracking research on domain-specific complex problem solving in interactive computer-based simulations was examined (Study 3). Twelve studies from different areas of VET were identified. The most common eye-tracking metrics are positional measures, such as the proportion of fixation time or total dwell time on AOIs. Counting measures often refer to the number of fixations and dwell times. Movement measures such as saccade directions or scan paths are rarely computed. Heatmaps and scanpaths are often compared qualitatively, with a notable lack of quantitative approaches to measure temporal patterns or similarities. Quantitative analysis applying t-tests and ANOVAs are often used for group comparisons (expert vs. novice, high vs. low performing groups). In addition, linear mixed effects models are useful to account for individual differences and increase statistical power. Post-hoc performance predictions using multimodal learning

analyses are already available in the literature, but real-time analysis of learning from eye-tracking in simulations is not yet practical. Combining eye-tracking with log file analysis, think-aloud protocols, and psychophysiological measures (e.g., EDA) can help to interpret ambiguous data and provide comprehensive insights into complex problem-solving processes in authentic computer-based simulations.

The next step was to use the findings from the review to conduct an initial eye-tracking study to investigate student performance and visual attention when using static and dynamic interactive graphics to solve economic tasks, using eye-tracking technology to observe learning behavior in a digital learning environment (Study 4). Analysis of performance differences showed that students performed better on economic learning tasks with dynamic, interactive graphs than with static graphs. This was evident on more difficult tasks. Visual attention was measured using eye-tracking metrics such as dwell time, fixation duration, and number of fixations. Students fixated significantly longer on dynamic graphs than on static graphs. However, overall fixation durations were short, suggesting rapid visual processing rather than deep cognitive engagement. Total fixation duration was a significant predictor of performance. Fixation duration predicted performance on tasks with dynamic but not static graphs, suggesting its importance in the study of learning processes with engaging material. This result suggests that dynamic graphs can better support the learning of students who are already familiar with static graphs. Dynamic, interactive graphs improve performance and engage visual attention more effectively than static graphs, especially for complex tasks.

In summary, the results of the five studies provide valuable insights into the impact of digitalization in business education. The first line of research identified the current use of digital tools in commercial schools, highlighting which digital tools are being integrated by participating teachers. The second line of research evaluated how to train prospective teachers to teach and learn with sophisticated digital tools (in this case an ERP system) and evaluated the reliable use of large open-source language models and discussed their potential in a two-way educational assessment workflow. The preparatory course for master's students to learn and teach business process knowledge with an advanced ERP system was then included in the program for student teachers at the University of Mannheim. Technological innovations such as large language models have also proven useful for a domain-specific assessment task in a two-way workflow. The third line of research analyzed individual learning processes in technology-rich environments. Eye-tracking was used as a methodological approach to investigate complex problem-solving processes in an authentic learning environment to determine differences in visual attention. After conducting a systematic review, the methodology was applied to understand learning

processes and improve instructional design in a computer-based learning environment for economic education.

8.2 Limitations and Future Research

Before drawing practical conclusions from the findings presented, this chapter discusses the limitations of the research and offers suggestions for future studies in the three main areas.

The current state of digital tools in commercial schools

Study 1 was conducted during the pandemic and thus has several limitations. The selected area and sample are not representative, and sample bias is possible because participation may have been skewed toward teachers interested in the topic. The results are based solely on teacher self-report (Chan, 2010) and should be further supported through objective data. The SAMR model to characterize teaching was challenging and yielded unsatisfactory results; a major limitation is that no conclusions can be drawn about the quality of teaching (Hamilton et al., 2016). Most of the digital tools used fall under the first two levels of the SAMR model. However, this does not indicate the quality of teaching or learning success, nor can it be determined whether the use of digital tools was a result of the pandemic or pre-existing practices. Further studies should link learning outcomes and teaching quality with digital tools, which was not possible during the pandemic, but would be an interesting research target.

Teaching and learning with sophisticated digital tools in VET

Study 2 represented a specific context (teaching business process knowledge within an advanced ERP system) and was conducted with a small sample size. The limitations of self-assessment should therefore be remedied by an additional, more objective data channel (e.g. by analyzing performance in the ERP system). In addition, social desirability could play a role due to the relationship between trainer and researcher (Nederhof, 1985). This suggests that future studies should not conduct authentic competence assessments with prospective teachers, but rather focus on real VET environments, e.g. assessing trainees' business process knowledge and ERP skills in vocational schools.

Study 5, an experiment with LLMs, has several limitations. The results relate to a single, dichotomous, and subjective rating task of classifying emails as polite or impolite. Tests with paid language models yielded only slightly better results. Therefore, free of charge and openly available models were used in the study to promote the democratization of AI and to encourage researchers to use open-source transformer models. At the same time, recent developments show that new models would be likely better able to solve the classification task (Schneider et

al., 2023). Future research should therefore focus on the investigation of polytomous variables with fewer subjective judgment components than in this experiment. At the same time, the reciprocal workflow between humans and AI must be considered as an independent subject of research (Ifenthaler & Schumacher, 2023; Molenaar, 2022).

Insight into learning processes in digital learning environments

Regarding study 3, publication bias is an important limitation of any literature review (Marks-Anglin & Chen, 2020). Although the review attempted to avoid this by using multiple databases, broad search terms, and reference checking, relevant literature may be missed. Generalizability is limited by the narrow range of domains and tasks, and there are biases in representation, with Western countries overrepresented. The limited number of studies analyzed is a major limitation of this review, which cannot provide general statistical results due to heterogeneity in sample size and study design. Future research should conduct a broader analysis and with sufficient number of findings, also meta-analyses, particularly on the relationship between performance differences and eye movement measures. Further studies should investigate problem-solving competences and learning-related stress (Kärner et al., 2018) and emotions using embedded experience sampling (Rausch et al., 2019) or facial expression recognition (Stöckli et al., 2018). By combining multiple data streams, e.g., from additional log files or retrospective think aloud, a multimodal analysis of learning could help to legitimize interpretations of ambiguous eye-tracking results. A research gap can be identified in eye-tracking studies in economic and business education.

Study 4 was conducted to fill this gap but comes with several limitations. The tasks were presented in a fixed (not randomized) order, with static graphs first and dynamic graphs second, which could influence the results. As a result, this study can only provide information about the added value of dynamic graphs on the performance of the students. The rating of tasks as easy or difficult was based on subjective ratings, but this was supported by solution rates. Although the study followed the standard recommendations, the classification parameters can significantly impact the selection of fixation candidates (Kosel et al., 2023). The duration of the fixations was very short. A future study must address trainees as test subjects (similar to Study 2) and introduce a randomized sequence. The study focused on retention and completion of single-choice tasks without testing transfer performance. A later post-test will be necessary to determine learning success. Financial compensation could also be helpful to keep the rates for no response low (funding approved, study in preparation).

8.3 Practical Implications and Conclusion

After discussing the limitations and potential future research directions for the five studies presented in this thesis, the dissertation will conclude with practical implementation suggestions based on the findings, as well as a summary of the overall relevance of this work.

Considering the current state of digital tools in vocational schools, the results suggest a potential for further transformation of teaching and learning in vocational schools in Germany. The current digital tools refer to the lower levels of the SAMR model, indicating that there might be some improvements for further digitalization in VET at the learning site of schools. However, these need to be implemented carefully, as Delcker and Ifenthaler suggest that coping strategies during the pandemic were supported by a well-defined school leadership agenda, a reliable technology infrastructure, and the willingness of teachers to adopt (2021). In addition, student teachers need to be led by example already at the university level, as it is assumed that digital tools will be used more intense and more successfully in VET if these educators have experienced digital tools as useful during their own training (Amhag et al., 2019). Such examples of good practice can be found in the use of ERP systems in vocational schools, as reported in Study 2.

With regard to teaching and learning with advanced digital tools in VET, prepared materials and learning environments supporting teaching, especially in this case by teachers for teachers, are helpful for teaching and learning with advanced digital tools such as ERP systems. Familiarization already at the university level might encounter the current lack of knowledge of (prospective) teachers among ERP systems as learning environments, which is probably one reason for the low level of integration of ERP systems in the education sector to date (Knigge et al., 2017). Combined with in-service training in the latter phases of teacher education, this could lead to an increased use of such learning environments in VET. However, the materials provided also have some limitations, as changes to the ERP tasks are likely to be difficult to make, as the ERP system provided only allows for a limited number of individual changes to teaching and learning. A more practical way would be to integrate AI tools to individualize and flexible teaching and learning, e.g. with a customized company per occupation. AI could help to individualize challenging tasks, industry-related providers and suppliers to consider the differences between professions. For example, a two-wheel mechatronics technician specializing in bicycle technology might be more attracted to the currently implemented standard company (Global Bikes) than a specialist drugstore salesperson.

As Study 5 shows, AI can be seen as a ‘game changer’ in education, assisting educators and learners in a variety of ways. Assisting with the task of grading may be one possibility that could soon become a reality. The reciprocal perspectives of AI and human intelligence in education would also be of great importance here (Ifenthaler & Schumacher, 2023), as machine scores indicate a certain probability of how reliably the result was determined. Low probability results could be filtered out and completed by humans, while supposedly easier high probability classifications could be automatically evaluated by machines, with humans giving AI more leeway. Prompt-based learning is a promising complement to human evaluators because it reduces the need for extensive training data, but still requires human supervision for reliability. Fine-tuned models offer higher accuracy but come at a higher cost due to manual data labeling. The integration of machine and human evaluation improves the quality of the evaluation and emphasizes the importance of a collaborative approach. Further research into prompter technology and the democratization of AI is essential to maximize the potential of these technologies in education. According to Ifenthaler et al. (2024), the three main challenges here are currently privacy and ethical use, trustworthy algorithms, and equity and fairness.

The technological advances also allow further conclusions to be drawn about the learning processes in digital learning environments in vocational education and training. A practical implication of the review (Study 3) is that the studies demonstrate the versatility of eye-tracking as a data collection method, as computer-based simulations are becoming increasingly important in VET and are suitable for measuring online learning. The results of study 4 show that visual attention differs depending on the type of stimuli (dynamic vs. static) and that instructional design can be tuned with information about visual attention. As a next step, a study will be conducted with students focusing on the micro-level of AOIs (see Ruf et al., 2023 for an overview) for diagrams, which is expected to provide further insights into the visual processing of diagrams in business classes. There is considerable potential for further research and integration of online and offline measures and multimodal data to improve the assessment and understanding of complex problem solving in computer-based learning environments in VET (e.g., Rausch et al., 2021).

In conclusion, the work presented provides a comprehensive overview of the impact of digitalization on VET with practical examples and conclusions for further research and practice. It emphasizes the importance of training prospective teachers in the use of digital tools and provides examples of the use of digital tools in AI and training on ERP systems. It also sheds light on online measurements by examining the current state of the art for eye-tracking research in computer-based simulations concerning VET and by conducting an experiment in a dedicated

economic education learning environment. As expected, digitalization will further influence VET research and practice now and in the future.

8.4 References

- Amhag, L., Hellström, L., & Stigmar, M. (2019). Teacher Educators' Use of Digital Tools and Needs for Digital Competence in Higher Education. *Journal of Digital Learning in Teacher Education*, 35(4), 203–220. <https://doi.org/10.1080/21532974.2019.1646169>
- Chan, D. (2010). So why ask me? Are self-report data really that bad? In R. J. Vandenberg & C. E. Lance (Eds.), *Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in Organizational and Social Sciences* (pp. 329–356). Routledge.
- Delcker, J., & Ifenthaler, D. (2021). Teachers' perspective on school development at German vocational schools during the Covid-19 pandemic. *Technology, Pedagogy and Education*, 30(1), 125–139. <https://doi.org/10.1080/1475939X.2020.1857826>
- Euler, D., & Wilbers, K. (2020). Berufsbildung in digitalen Lernumgebungen [Vocational training in digital learning environments]. In R. Arnold, A. Lipsmeier, & M. Rohs (Eds.), *Handbuch Berufsbildung* (pp. 427–438). Springer. https://doi.org/10.1007/978-3-658-19312-6_34
- Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The Substitution Augmentation Modification Redefinition (SAMR) Model: A Critical Review and Suggestions for its Use. *TechTrends*, 60(5), 433–441. <https://doi.org/10.1007/s11528-016-0091-y>
- Ifenthaler, D., Majumdar, R., Gorissen, P., Judge, M., Mishra, S., Raffaghelli, J., & Shimada, A. (2024). Artificial Intelligence in Education: Implications for Policymakers, Researchers, and Practitioners. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-024-09747-0>
- Ifenthaler, D., & Schumacher, C. (2023). Reciprocal issues of artificial and human intelligence in education. *Journal of Research on Technology in Education*, 55(1), 1–6. <https://doi.org/10.1080/15391523.2022.2154511>
- Kärner, T., Minkley, N., Rausch, A., Schley, T., & Sembill, D. (2018). Stress and Resources in Vocational Problem Solving. *Vocations and Learning*, 11(3), 365–398. <https://doi.org/10.1007/s12186-017-9193-8>
- Knigge, M., Prifti, L., Kienegger, H., & Krcmar, H. (2017). Teaching enterprise organization and enterprise resource planning systems in schools: Playing a serious game with pupils. *2017 IEEE Global Engineering Education Conference (EDUCON)*, 486–495. <https://doi.org/10.1109/EDUCON.2017.7942891>
- Kosel, C., Voggenreiter, A., Pfeffer, J., & Seidel, T. (2023). Measuring Teachers' Visual Expertise Using the Gaze Relational Index Based on Real-World Eyetracking Data and Varying Velocity Thresholds. *Journal of Expertise*, 6(2), 267–281.
- Marks-Anglin, A., & Chen, Y. (2020). A historical review of publication bias. *Research Synthesis Methods*, 11(6), 725–742. <https://doi.org/10.1002/jrsm.1452>
- Mayer, C. W., Findeisen, S., Guggemos, J., & Seifried, J. (under review). Visual attention while solving economic learning tasks using dynamic graphs – An eye-tracking study. *Journal of Economic Education*.
- Mayer, C. W., Gentner, S., & Seifried, J. (2023). Digitale Unterrichtspraxis an kaufmännischen Schulen in der Corona-Pandemie—Eine Momentaufnahme [Digital teaching practice at

- commercial schools during the corona pandemic - a snapshot]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 119(3), 531–552. <https://doi.org/10.25162/zbw-2023-0021>
- Mayer, C. W., Ludwig, S., & Brandt, S. (2022). Prompt text classifications with transformer models! An exemplary introduction to prompt-based learning with large language models. *Journal of Research on Technology in Education*, 55(1), 1–17. <https://doi.org/10.1080/15391523.2022.2142872>
- Mayer, C. W., Rausch, A., & Seifried, J. (2023). Analysing domain-specific problem-solving processes within authentic computer-based learning and training environments by using eye-tracking: A scoping review. *Empirical Research in Vocational Education and Training*, 15(1), 2. <https://doi.org/10.1186/s40461-023-00140-2>
- Mayer, C. W., & Seifried, J. (2024). Looking behind the scenes—Being able to master ERP systems as a goal of vocational education and training. In Cedefop & OECD (Eds.), *Apprenticeships and the digital transition* (pp. 118–127). Publications Office.
- Molenaar, I. (2022). The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. *Computers and Education: Artificial Intelligence*, 3, 100070. <https://doi.org/10.1016/j.caeai.2022.100070>
- Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3), 263–280. <https://doi.org/10.1002/ejsp.2420150303>
- Rausch, A., Deutscher, V., Seifried, J., Brandt, S., & Winther, E. (2021). Die web-basierte Bürosimulation LUCA – Funktionen, Einsatzmöglichkeiten und Forschungsausblick [The web-based office simulation LUCA - functions, possible applications and research outlook]. *Zeitschrift für Beruf und Wirtschaftspädagogik*, 117(3), 372–394. <https://doi.org/10.25162/zbw-2021-0017>
- Rausch, A., Kögler, K., & Seifried, J. (2019). Validation of Embedded Experience Sampling (EES) for Measuring Non-cognitive Facets of Problem-Solving Competence in Scenario-Based Assessments. *Frontiers in Psychology*, 10, 1200. <https://doi.org/10.3389/fpsyg.2019.01200>
- Ruf, V., Horrer, A., Berndt, M., Hofer, S. I., Fischer, F., Fischer, M. R., Zottmann, J. M., Kuhn, J., & Küchemann, S. (2023). A Literature Review Comparing Experts' and Non-Experts' Visual Processing of Graphs during Problem-Solving and Learning. *Education Sciences*, 13(2), 216. <https://doi.org/10.3390/educsci13020216>
- Schneider, J., Schenk, B., & Niklaus, C. (2023). *Towards LLM-based Autograding for Short Textual Answers*. <https://doi.org/10.48550/ARXIV.2309.11508>
- Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2018). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods*, 50(4), 1446–1460. <https://doi.org/10.3758/s13428-017-0996-1>
- Wuttke, E., Seifried, J., & Niegemann, H. M. (Eds.). (2020). *Vocational education and training in the age of digitization: Challenges and opportunities*. Barbara Budrich.

Complete List of References

- Abele, S. (2018). Diagnostic Problem-Solving Process in Professional Contexts: Theory and Empirical Investigation in the Context of Car Mechatronics Using Computer-Generated Log-Files. *Vocations and Learning*, *11*(1), 133–159. <https://doi.org/10.1007/s12186-017-9183-x>
- Abele, S., Ostertag, R., Peissner, M., & Schuller, A. (2017). Eine Eye-Tracking-Studie zum diagnostischen Problemlöseprozess. Bedeutung der Informationsrepräsentation für den diagnostischen Problemlöseerfolg [An eye-tracking study on the diagnostic problem-solving process. The importance of information representation for diagnostic problem-solving success]. *Zeitschrift Für Berufs- Und Wirtschaftspädagogik*, *113*(1), 86–109.
- Abid, A., Farooqi, M., & Zou, J. (2021). Persistent anti-muslim bias in large language models. *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 298–306.
- Achtenhagen, F. (2012). The curriculum-instruction-assessment triad. *Empirical Research in Vocational Education and Training*, *4*(1), 5–25.
- Ahmed, A., Hurwitz, D., Gestson, S., & Brown, S. (2021). Differences between Professionals and Students in Their Visual Attention on Multiple Representation Types While Solving an Open-Ended Engineering Design Problem. *Journal of Civil Engineering Education*, *147*(3), 04021005. [https://doi.org/10.1061/\(ASCE\)EI.2643-9115.0000044](https://doi.org/10.1061/(ASCE)EI.2643-9115.0000044)
- Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, *125*, 413–428. <https://doi.org/10.1016/j.compedu.2018.06.023>
- Amhag, L., Hellström, L., & Stigmar, M. (2019). Teacher Educators' Use of Digital Tools and Needs for Digital Competence in Higher Education. *Journal of Digital Learning in Teacher Education*, *35*(4), 203–220. <https://doi.org/10.1080/21532974.2019.1646169>
- Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes: Limits of the eye-mind hypothesis. *Psychological Science*, *15*(4), 225–231.
- Andrade, H. L. (2019). A Critical Review of Research on Student Self-Assessment. *Frontiers in Education*, *4*, 87. <https://doi.org/10.3389/educ.2019.00087>
- Arksey, H., & O'Malley, L. (2005). Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19–32. <https://doi.org/10.1080/1364557032000119616>
- Ashraf, H., Sodergren, M. H., Merali, N., Mylonas, G., Singh, H., & Darzi, A. (2018). Eye-tracking technology in medical education: A systematic review. *Medical Teacher*, *40*(1), 62–69. <https://doi.org/10.1080/0142159X.2017.1391373>
- Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V. 2. *The Journal of Technology, Learning and Assessment*, *4*(3).
- Autor:innengruppe Bildungsberichterstattung. (2024). *Bildung in Deutschland 2024. Ein indikatoren-gestützter Bericht mit einer Analyse zu beruflicher Bildung [Education in Germany 2024. An indicator-based report with an analysis of vocational education and training]*. wbv.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- Baddeley, A. D. (1997). *Human memory: Theory and practice*. psychology press.

- Beck, K., Landenberger, M., & Oser, F. (Eds.). (2016). *Technologiebasierte Kompetenzmessung in der beruflichen Bildung: Ergebnisse aus der BMBF-Förderinitiative ASCOT [Technology-based skills assessment in vocational education and training: results from the BMBF funding initiative ASCOT]*. W. Bertelsmann.
- Becker, A. (2020). Post-collection Data Capture, Scoring, and Processing. In H. Wagemaker (Ed.), *Reliability and Validity of International Large-Scale Assessment* (pp. 151–167). Springer. https://doi.org/10.1007/978-3-030-53081-5_9
- Bennett, S., Maton, K., & Kervin, L. (2008). The ‘digital natives’ debate: A critical review of the evidence. *British Journal of Educational Technology*, 39(5), 775–786. <https://doi.org/10.1111/j.1467-8535.2007.00793.x>
- Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, 26(1), 87–122. <https://doi.org/10.1007/s12528-013-9077-3>
- Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education*, 101, 150–167. <https://doi.org/10.1016/j.compedu.2016.06.005>
- Beseiso, M., Alzubi, O. A., & Rashaideh, H. (2021). A novel automated essay scoring approach for reliable higher educational assessments. *Journal of Computing in Higher Education*, 33(3), 727–746. <https://doi.org/10.1007/s12528-021-09283-1>
- Billett, S. (2021). Mediating worklife learning and the digitalisation of work. *British Journal of Educational Technology*, 52(4), 1580–1593. <https://doi.org/10.1111/bjet.13115>
- Bin, L., & Jian-Min, Y. (2011). Automated Essay Scoring Using Multi-classifier Fusion. In Y. Wu (Ed.), *Computing and Intelligent Systems* (pp. 151–157). Springer. https://doi.org/10.1007/978-3-642-24010-2_21
- BLV (Verband der Lehrerinnen und Lehrer an beruflichen Schulen in Baden-Württemberg). (2021). *Ergebnisse der BLV-Umfrage zu Fernunterricht [Results of the BLV survey on distance learning]*. Berufsschullehrerverband Baden-Württemberg. <https://blv-bw.de/ergebnisse-der-blv-umfrage-zum-fernunterricht-16306/>
- Bos, W., Eickelmann, B., Gerick, J., Goldhammer, F., Schaumburg, H., Schwippert, K., Senkbeil, M., Schulz-Zander, R., & Wendt, H. (2014). *ICILS 2013: Computer-und informationsbezogene Kompetenzen von Schülerinnen und Schülern in der 8. Jahrgangsstufe im internationalen Vergleich [ICILS 2013: International comparison of computer and information-related skills of pupils in year 8]*. Waxmann.
- Bostan, L.-A.-M., & Klinger, R. (2018). An Analysis of Annotated Corpora for Emotion Classification in Text. *Proceedings of the 27th International Conference on Computational Linguistics*, 2104–2119. <https://aclanthology.org/C18-1179>
- Brandt, S., Rausch, A., & Kögler, K. (2016). *A Scoring Procedure for Complex Assessments Focusing on Validity and Appropriate Reliability*. 1–20.
- Breckwoldt, J., Gruber, H., & Wittmann, A. (2014). Simulation Learning. In S. Billett, C. Harteis, & H. Gruber (Eds.), *International Handbook of Research in Professional and Practice-based Learning* (pp. 673–698). Springer. https://doi.org/10.1007/978-94-017-8902-8_25
- Brehmer, B., & Dörner, D. (1993). Experiments with computer-simulated microworlds: Escaping both the narrow straits of the laboratory and the deep blue sea of the field study. *Computers in Human Behavior*, 9(2–3), 171–184. [https://doi.org/10.1016/0747-5632\(93\)90005-D](https://doi.org/10.1016/0747-5632(93)90005-D)

- Brinda, T., Brüggem, N., Diethelm, I., Knaus, T., Kommer, S., Kopf, C., Missomelius, P., Leschke, R., Tilemann, F., & Weich, A. (2019). Frankfurt-Dreieck zur Bildung in der digital vernetzten Welt: Ein interdisziplinäres Modell [Frankfurt triangle for education in the digitally networked world: an interdisciplinary model]. *18. GI-Fachtagung Informatik und Schule - Informatik für alle*, 25–33. <https://doi.org/10.18420/INFOS2019-A1>
- Brinda, T., Diethelm, I., Gemulla, R., Romeike, R., Schöning, J., Schulte, C., Bartoschek, T., Bergner, N., Dietrich, L., Döbeli, B., Fries, R., Hellmig, L., Herzig, B., Hollatz, J., Jörisen, B., Kommer, S., Mittag, A., Kusterer, P., Oberweis, A., ... Zimnol, M. (2016). *Dagstuhl-Erklärung: Bildung in der digitalen vernetzten Welt*. Gesellschaft für Informatik e.V. <https://dl.gi.de/handle/20.500.12116/44026>
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., & Askell, A. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
- Brückner, S., Zlatkin-Troitschanskaia, O., Küchemann, S., Klein, P., & Kuhn, J. (2020). Changes in Students' Understanding of and Visual Attention on Digitally Represented Graphs Across Two Domains in Higher Education: A Postreplication Study. *Frontiers in Psychology*, 11, 2090. <https://doi.org/10.3389/fpsyg.2020.02090>
- Brynjolfsson, E., & McAfee, A. (2016). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton.
- Buettner, R., Sauer, S., Maier, C., & Eckhardt, A. (2018). Real-time Prediction of User Performance based on Pupillary Assessment via Eye-Tracking. *AIS Transactions on Human-Computer Interaction*, 10(1), 26–60. <https://doi.org/10.17705/1thci.00103>
- Bujard, M., von den Driesch, E., Ruckdeschel, K., Lass, I., Thönnissen, C., Schumann, A., & Schneider, N. F. (2021). *Belastungen von Kindern, Jugendlichen und Eltern in der Corona-Pandemie [Stresses on children, young people and parents during the coronavirus pandemic]*. (2/2021; BiB.Bevölkerungs.Studien). Bundesinstitut für Bevölkerungsforschung. <https://doi.org/10.12765/bro-2021-02>
- Bundesinstitut für Berufsbildung. (2022, August 17). *Digitale Transformationen [Digital transformations]*. <https://www.bibb.de/de/120255.php>
- Bürger, N., Haselmann, S., Baumgart, J., Prinz, G., Girnat, B., Meisert, A., Menthe, J., Schmidt-Thieme, B., & Wecker, C. (2021). Jenseits von Professionswissen: Eine systematische Überblicksarbeit zu einstellungs- und motivationsbezogenen Einflussfaktoren auf die Nutzung digitaler Technologien im Unterricht [Beyond professional knowledge: A systematic review of attitudinal and motivational factors influencing the use of digital technologies in the classroom]. *Zeitschrift für Erziehungswissenschaft*, 24(5), 1087–1112. <https://doi.org/10.1007/s11618-021-01050-3>
- Busch, M. (2020). *55 Webtools für den Unterricht: Einfach, konkret, step by step [55 web tools for teaching: simple, concrete, step by step]* (4. komplett überarbeitete und aktualisierte Auflage). Auer.
- Carter, B. T., & Luke, S. G. (2020). Best practices in eye tracking research. *International Journal of Psychophysiology*, 155, 49–62.
- Catrysse, L., Gijbels, D., Donche, V., De Maeyer, S., Lesterhuis, M., & Van den Bossche, P. (2018). How are learning strategies reflected in the eyes? Combining results from self-reports and eye-tracking. *British Journal of Educational Psychology*, 88(1), 118–137. <https://doi.org/10.1111/bjep.12181>
- Causse, M., Lancelot, F., Maillant, J., Behren, J., Cousy, M., & Schneider, N. (2019). Encoding decisions and expertise in the operator's eyes: Using eye-tracking as input for system

- adaptation. *International Journal of Human-Computer Studies*, 125, 55–65. <https://doi.org/10.1016/j.ijhcs.2018.12.010>
- Celik, I., Dindar, M., Muukkonen, H., & Järvelä, S. (2022). The Promises and Challenges of Artificial Intelligence for Teachers: A Systematic Review of Research. *TechTrends*, 66(4), 616–630. <https://doi.org/10.1007/s11528-022-00715-y>
- Cerri, S. A. (2012). Advanced Learning Technologies. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 154–157). Springer. https://doi.org/10.1007/978-1-4419-1428-6_57
- Chan, B., Schweter, S., & Möller, T. (2020). German's Next Language Model. *Proceedings of the 28th International Conference on Computational Linguistics*, 6788–6796.
- Chan, D. (2010). So why ask me? Are self-report data really that bad? In R. J. Vandenberg & C. E. Lance (Eds.), *Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in Organizational and Social Sciences* (pp. 329–356). Routledge.
- Chaudhury, P. (2023). Asynchronous learning design—Lessons for the post-pandemic world of higher education. *The Journal of Economic Education*, 54(2), 214–223. <https://doi.org/10.1080/00220485.2023.2174233>
- Chauhan, S. (2017). A meta-analysis of the impact of technology on learning effectiveness of elementary students. *Computers & Education*, 105, 14–30. <https://doi.org/10.1016/j.compedu.2016.11.005>
- Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-Based Learning in Higher Education: A Meta-Analysis. *Review of Educational Research*, 90(4), 499–541. <https://doi.org/10.3102/0034654320933544>
- Clarebout, G., Elen, J., Lowyck, J., Van den Ende, J., & Van den Enden, E. (2009). Tropical Medicine Open Learning Environment: In P. L. Rogers, G. A. Berg, J. V. Boettcher, C. Howard, L. Justice, & K. D. Schenk (Eds.), *Encyclopedia of Distance Learning, Second Edition* (pp. 2155–2159). IGI Global. <https://doi.org/10.4018/978-1-60566-198-8.ch318>
- Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (Fourth edition). Wiley.
- Cloude, E. B., Azevedo, R., Winne, P. H., Biswas, G., & Jang, E. E. (2022). System design for using multimodal trace data in modeling self-regulated learning. *Frontiers in Education*, 7, 928632. <https://doi.org/10.3389/educ.2022.928632>
- Cloude, E. B., Dever, D. A., Wiedbusch, M. D., & Azevedo, R. (2020). Quantifying Scientific Thinking Using Multichannel Data With Crystal Island: Implications for Individualized Game-Learning Analytics. *Frontiers in Education*, 5, 572546. <https://doi.org/10.3389/educ.2020.572546>
- Coban, M., Bolat, Y. I., & Goksu, I. (2022). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review*, 36, 100452. <https://doi.org/10.1016/j.edurev.2022.100452>
- Cochran, K. F., DeRuiter, J. A., & King, R. A. (1993). Pedagogical Content Knowing: An Integrative Model for Teacher Preparation. *Journal of Teacher Education*, 44(4), 263–272. <https://doi.org/10.1177/0022487193044004004>
- Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220. <https://doi.org/10.1037/h0026256>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Academic press.

- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Cohn, E., & Cohn, S. (1994). Graphs and learning in principles of economics. *The American Economic Review*, *84*(2), 197–200.
- Cohn, E., Cohn, S., Balch, D. C., & Bradley, J. (2001). Do Graphs Promote Learning in Principles of Economics? *The Journal of Economic Education*, *32*(4), 299–310. <https://doi.org/10.1080/00220480109596110>
- Cohn, E., Cohn, S., Balch, D. C., & Bradley, J. (2004). The Relation between Student Attitudes toward Graphs and Performance in Economics. *The American Economist*, *48*(2), 41–52. <https://doi.org/10.1177/056943450404800203>
- Conati, C., Jaques, N., & Muir, M. (2013). Understanding Attention to Adaptive Hints in Educational Games: An Eye-Tracking Study. *International Journal of Artificial Intelligence in Education*, *23*, 136–161. <https://doi.org/10.1007/s40593-013-0002-8>
- Cook, D. A., Hamstra, S. J., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., & Hatala, R. (2013). Comparative effectiveness of instructional design features in simulation-based education: Systematic review and meta-analysis. *Medical Teacher*, *35*(1), e867–e898. <https://doi.org/10.3109/0142159X.2012.714886>
- Coskun, A., & Cagiltay, K. (2022). A systematic review of eye-tracking-based research on animated multimedia learning. *Journal of Computer Assisted Learning*, *38*(2), 581–598. <https://doi.org/10.1111/jcal.12629>
- Cuve, H. C., Stojanov, J., Roberts-Gaal, X., Catmur, C., & Bird, G. (2022). Validation of Gaze-point low-cost eye-tracking and psychophysiology bundle. *Behavior Research Methods*, *54*(2), 1027–1049. <https://doi.org/10.3758/s13428-021-01654-x>
- Deissinger, T. (2015). The German dual vocational education and training system as ‘good practice’? *Local Economy: The Journal of the Local Economy Policy Unit*, *30*(5), 557–567. <https://doi.org/10.1177/0269094215589311>
- Delcker, J., & Ifenthaler, D. (2021). Teachers’ perspective on school development at German vocational schools during the Covid-19 pandemic. *Technology, Pedagogy and Education*, *30*(1), 125–139. <https://doi.org/10.1080/1475939X.2020.1857826>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. <https://doi.org/10.48550/ARXIV.1810.04805>
- Dewan, M. A. A., Murshed, M., & Lin, F. (2019). Engagement detection in online learning: A review. *Smart Learning Environments*, *6*, 1–20. <https://doi.org/10.1186/s40561-018-0080-z>
- Dilger, B. (2021). Lernen aus den Erfahrungen mit Emergency-Remote-Lehre aus Studierenden- und Dozierenden-Perspektive [Learning from experiences with emergency remote teaching from the perspective of students and lecturers]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, *117*(1), 3–13. <https://doi.org/10.25162/zbw-2021-0001>
- Dillenbourg, P. (2013). Design for classroom orchestration. *Computers & Education*, *69*, 485–492. <https://doi.org/10.1016/j.compedu.2013.04.013>
- Di Stasi, L., Antoli, A., & Canas, J. (2011). Main sequence: An index for detecting mental workload variation in complex tasks. *Applied Ergonomics*, *42*(6), 807–813. <https://doi.org/10.1016/j.apergo.2011.01.003>

- Drossel, K., Eickelmann, B., Schaumburg, H., & Labusch, A. (2019). Nutzung digitaler Medien und Prädiktoren aus der Perspektive der Lehrerinnen und Lehrer im internationalen Vergleich [Use of digital media and predictors from the perspective of teachers in an international comparison]. In B. Eickelmann, W. Bos, J. Gerick, F. Goldhammer, H. Schaumburg, K. Schwippert, M. Senkbeil, & J. Vahrenhold (Hrsg.), *ICILS 2018 #Deutschland. Computer- und informationsbezogene Kompetenzen von Schülerinnen und Schülern im zweiten internationalen Vergleich und Kompetenzen im Bereich Computational Thinking* (S. 205–240). Waxmann. <https://doi.org/10.25656/01:18325>
- Drugova, E., Zhuravleva, I., Aiusheeva, M., & Grits, D. (2021). Toward a model of learning innovation integration: TPACK-SAMR based analysis of the introduction of a digital learning environment in three Russian universities. *Education and Information Technologies*, 26(4), 4925–4942. <https://doi.org/10.1007/s10639-021-10514-2>
- Döbeli Honegger, B. (2021). Covid-19 und die digitale Transformation in der Schweizer Lehrerinnen- und Lehrerbildung. *Beiträge zur Lehrerinnen- und Lehrerbildung*, 39(3), 411–422. <https://doi.org/10.25656/01:23693>
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften [Research methods and evaluation in the social and human sciences]* (5. vollständig überarbeitete, aktualisierte und erweiterte Auflage). Springer.
- Dörner, D. (1987). *Problemlösen als Informationsverarbeitung [Problem solving as information processing]*. Kohlhammer.
- Dörner, D. (1997). *The logic of failure: Recognizing and avoiding error in complex situations*. Basic Books.
- Dörner, D., & Funke, J. (2017). Complex problem solving: What it is and what it is not. *Frontiers in Psychology*, 8, 1153. <https://doi.org/10.3389/fpsyg.2017.01153>
- Dubovi, I. (2022). Cognitive and emotional engagement while learning with VR: The perspective of multimodal methodology. *Computers & Education*, 183, 104495. <https://doi.org/10.1016/j.compedu.2022.104495>
- Duchowski, A. T. (2017). *Eye Tracking Methodology*. Springer. <https://doi.org/10.1007/978-3-319-57883-5>
- Duncker, K. (1945). On problem-solving. (L. S. Lees, Trans.). *Psychological Monographs*, 58(5), i–113. <https://doi.org/10.1037/h0093599>
- Economou, A. (2023). *SELFIEforTEACHERS. Designing and developing a self-reflection tool for teachers' digital competence*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2760/40528>
- Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, 1(1), 56–75.
- Emerson, A., Cloude, E. B., Azevedo, R., & Lester, J. (2020). Multimodal learning analytics for game-based learning. *British Journal of Educational Technology*, 51(5), 1505–1526. <https://doi.org/10.1111/bjet.12992>
- Esau, T., & Fletcher, S. (2018). Prozessorientierte Analyse von konstruktiven Problemlöseprozessen auf Basis von Eye-Tracking-Aufnahmen [Process-oriented analysis of constructive problem-solving processes on the basis of eye-tracking recordings]. *Journal of Technical Education*, 6(1), 2198–0306. <https://doi.org/10.48513/joted.v6i1.116>
- Euler, D., & Wilbers, K. (2020). Berufsbildung in digitalen Lernumgebungen [Vocational training in digital learning environments]. In R. Arnold, A. Lipsmeier, & M. Rohs (Eds.),

- Handbuch Berufsbildung* (pp. 427–438). Springer. https://doi.org/10.1007/978-3-658-19312-6_34
- European Centre for the Development of Vocational Training. (2022). *Setting Europe on course for a human digital transition: New evidence from Cedefop's second European skills and jobs survey*. Publications Office. <https://data.europa.eu/doi/10.2801/253954>
- European Commission (Ed.). (2023). *Digital Economy and Society Index (DESI) 2023: Integration of digital technology*. <https://digital-decade-desi.digital-strategy.ec.europa.eu/datasets/desi/charts>
- Evans, C., & Robertson, W. (2020). The four phases of the digital natives debate. *Human Behavior and Emerging Technologies*, 2(3), 269–277. <https://doi.org/10.1002/hbe2.196>
- Fernandez, C. (2014). Knowledge Base for Teaching and Pedagogical Content Knowledge (PCK): Some Useful Models and Implications for Teachers' Training. *Problems of Education in the 21st Century*, 60(1), 79–100. <https://doi.org/10.33225/pec/14.60.79>
- Findeisen, S., Messner, J., & Guggemos, J. (2022). Dynamische, interaktive Visualisierungen – Erkenntnisse eines systematischen Reviews und Vorstellung einer digitalen Lernumgebung für die Domäne Wirtschaft [Dynamic, interactive visualizations—Findings of a systematic review and presentation of a digital learning environment for the business domain]. In K. Kögler, U. Weyland, & H.-H. Kremer (Eds.), *Jahrbuch der berufs- und wirtschaftspädagogischen Forschung 2022* (pp. 219–235). Barbara Budrich. <https://doi.org/10.25656/01:26524>
- Fischer, A., & Neubert, J. C. (2015). The multiple faces of complex problems: A model of problem solving competency and its implications for training and assessment. *Journal of Dynamic Decision Making*, 1, 1–13. <https://doi.org/10.11588/jddm.2015.1.23945>
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2019). *How to design and evaluate research in education*. McGraw Hill.
- Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2019). *Preparing for life in a digital world: IEA international computer and information literacy study 2018 international report*. International Association for the Evaluation of Educational Achievement (IEA). <https://www.iea.nl/studies/iea/icils/2018>
- Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences*, 116(14), 6531–6539. <https://doi.org/10.1073/pnas.1900949116>
- Frensch, P. A., & Funke, J. (1995). Definitions, traditions, and a general framework for understanding complex problem solving. In P. A. Frensch & J. Funke (Eds.), *Complex problem solving. The European perspective* (pp. 3–25). Lawrence Erlbaum.
- Frenzel-Piasentin, A., Muench, J., Bruckner, M., & Veit, D. (2021). Digitization or Digitalization? – Toward an Understanding of Definitions, Use and Application in IS Research. *AMCIS 2021 Proceedings*, 18. https://aisel.aisnet.org/amcis2021/adv_info_systems_general_track/adv_info_systems_general_track/18
- Friebel-Piechotta, S. (2022). Economic models in German textbooks. *JSSE - Journal of Social Science Education*, 21(2), 82–113. <https://doi.org/10.11576/JSSE-4854>
- Funke, J. (2012). Complex Problem Solving. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 682–685). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_685

- Gabriel, F., Cloude, E. B., & Azevedo, R. (2022). Using Learning Analytics to Measure Motivational and Affective Processes During Self-Regulated Learning with Advanced Learning Technologies. In Y. “Elle” Wang, S. Joksimović, M. O. Z. San Pedro, J. D. Way, & J. Whitmer (Eds.), *Social and Emotional Learning and Complex Skills Assessment* (pp. 93–108). Springer. https://doi.org/10.1007/978-3-031-06333-6_6
- Gal, P., Nicoletti, G., Renault, T., Sorbe, S., & Timiliotis, C. (2019). *Digitalisation and productivity: In search of the holy grail – Firm-level empirical evidence from EU countries* (OECD Economics Department Working Papers 1533). OECD Publishing. <https://doi.org/10.1787/5080f4b6-en>
- Gao, L., Biderman, S., Black, S., Golding, L., Hoppe, T., Foster, C., Phang, J., He, H., Thite, A., & Nabeshima, N. (2020). The pile: An 800gb dataset of diverse text for language modeling. *arXiv Preprint arXiv:2101.00027*.
- Gao, T., Fisch, A., & Chen, D. (2021). Making Pre-trained Language Models Better Few-shot Learners. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, 3816–3830.
- Gegenfurtner, A., Lehtinen, E., & Säljö, R. (2011). Expertise Differences in the Comprehension of Visualizations: A Meta-Analysis of Eye-Tracking Research in Professional Domains. *Educational Psychology Review*, 23(4), 523–552. <https://doi.org/10.1007/s10648-011-9174-7>
- Gegenfurtner, A., & Seppänen, M. (2013). Transfer of expertise: An eye tracking and think aloud study using dynamic medical visualizations. *Computers & Education*, 63, 393–403. <https://doi.org/10.1016/j.compedu.2012.12.021>
- Gess-Newsome, J. (1999). Pedagogical Content Knowledge: An Introduction and Orientation. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining Pedagogical Content Knowledge* (pp. 3–17). Kluwer Academic Publishers. https://doi.org/10.1007/0-306-47217-1_1
- Gillioz, A., Casas, J., Mugellini, E., & Khaled, O. A. (2020). Overview of the Transformer-based Models for NLP Tasks. *15th Conference on Computer Science and Information Systems (FedCSIS)*, 179–183. <https://doi.org/10.15439/2020F20>
- Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (2014). The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment. *Journal of Educational Psychology*, 106(3), 608–626. <https://doi.org/10.1037/a0034716>
- Gomes, J. S., Yassine, M., Worsley, M., & Blikstein, P. (2013). Analysing Engineering Expertise of High School Students Using Eye Tracking and Multimodal Learning Analytics. In S. D’Mello, R. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (pp. 1–3).
- Gorshid, G. D., Mayer, C., Rausch, A., & Seifried, J. (2022). Das LUCA-Dashboard im Usability-Test-Eine gaze-cued retrospective Think-Aloud-Studie [The LUCA dashboard in a usability test—A retrospective think-aloud study]. In S. Schumann, S. Seeber, & S. Abele (Eds.), *Digitale Transformation in der Berufsbildung* (pp. 189–212). wbv.
- Gouseti, A. (2010). Web 2.0 and education: Not just another case of hype, hope and disappointment? *Learning, Media and Technology*, 35(3), 351–356. <https://doi.org/10.1080/17439884.2010.509353>

- Gradillas, M., & Thomas, L. D. W. (2023). Distinguishing digitization and digitalization: A systematic review and conceptual framework. *Journal of Product Innovation Management, Early View*, 1–32. <https://doi.org/10.1111/jpim.12690>
- Grossman, P. L. (1990). *The making of a teacher: Teacher knowledge and teacher education*. Teachers College Press.
- Grub, A.-S., Biermann, A., Lewalter, D., & Brünken, R. (2022). Professional Vision and the Compensatory Effect of a Minimal Instructional Intervention: A Quasi-Experimental Eye-Tracking Study With Novice and Expert Teachers. *Frontiers in Education, 7*, 890690. <https://doi.org/10.3389/educ.2022.890690>
- Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics, 11*(1), 1–21. <https://doi.org/10.1080/00401706.1969.10490657>
- Guan, C., Mou, J., & Jiang, Z. (2020). Artificial intelligence innovation in education: A twenty-year data-driven historical analysis. *International Journal of Innovation Studies, 4*(4), 134–147. <https://doi.org/10.1016/j.ijis.2020.09.001>
- Haider, H., & Frensch, P. A. (1996). The role of information reduction in skill acquisition. *Cognitive Psychology, 30*(3), 304–337.
- Haider, H., & Frensch, P. A. (1999). Eye movement during skill acquisition: More evidence for the information-reduction hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(1), 172–190. <https://doi.org/10.1037/0278-7393.25.1.172>
- Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The Substitution Augmentation Modification Redefinition (SAMR) Model: A Critical Review and Suggestions for its Use. *TechTrends, 60*(5), 433–441. <https://doi.org/10.1007/s11528-016-0091-y>
- Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., & Xu, Y. (2022). A survey on vision transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 45*(1), 87–110.
- Hannafin, M. J. (1995). Open-Ended Learning Environments: Foundations, Assumptions, and Implications for Automated Design. In R. D. Tennyson & A. E. Barron (Eds.), *Automating Instructional Design: Computer-Based Development and Delivery Tools* (pp. 101–129). Springer. https://doi.org/10.1007/978-3-642-57821-2_5
- Happ, R., Förster, M., Zlatkin-Troitschanskaia, O., & Carstensen, V. (2016). Assessing the previous economic knowledge of beginning students in Germany: Implications for teaching economics in basic courses. *Citizenship, Social and Economics Education, 15*(1), 45–57. <https://doi.org/10.1177/2047173416646597>
- Happ, R., Zlatkin-Troitschanskaia, O., & Schmidt, S. (2016). An analysis of economic learning among undergraduates in introductory economics courses in Germany. *The Journal of Economic Education, 47*(4), 300–310. <https://doi.org/10.1080/00220485.2016.1213686>
- Harder, A., Imboden, S., Glassey-Previdoli, D., & Schumann, S. (2020). Schulleitungshandeln in Zeiten der digitalen Transformation: „Business as usual“ oder „Alles ist neu“? [School management in times of digital transformation: "Business as usual" or "Everything is new"?] *bwp@, Profil 6: Digitale Festschrift für EVELINE WUTTKE*, 18.
- Helm, C., Huber, S., & Loisinger, T. (2021). Was wissen wir über schulische Lehr-Lern-Prozesse im Distanzunterricht während der Corona-Pandemie? – Evidenz aus Deutschland, Österreich und der Schweiz [What do we know about school teaching-learning processes in distance learning during the coronavirus pandemic? - Evidence from Germany, Austria and Switzerland]. *Zeitschrift für Erziehungswissenschaft, 24*(2), 237–311. <https://doi.org/10.1007/s11618-021-01000-z>

- Helsper, E. J., & Eynon, R. (2010). Digital natives: Where is the evidence? *British Educational Research Journal*, 36(3), 503–520. <https://doi.org/10.1080/01411920902989227>
- Herzig, B. (2007). Medien in komplexen Lernumgebungen [Media in complex learning environments]. In H.-H. Kremer (Hrsg.), *Paderborner Forschungs- und Entwicklungswerkstatt: Forschungsfragen und -konzepte der beruflichen Bildung* (Bd. 1, S. 157–174). Eusl.
- Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. Wiley.
- Hill, C. D., & Stegner, T. (2003). Which Students Benefit from Graphs in a Principles of Economics Class? *The American Economist*, 47(2), 69–77. <https://doi.org/10.1177/056943450304700206>
- Hillmayr, D., Ziernwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897. <https://doi.org/10.1016/j.compedu.2020.103897>
- Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6(2), 107–116.
- Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hodges, C., Moore, S., Locke, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *Educause review*, 27(1), 1–9.
- Höffler, T. N. (2010). Spatial Ability: Its Influence on Learning with Visualizations—a Meta-Analytic Review. *Educational Psychology Review*, 22(3), 245–269. <https://doi.org/10.1007/s10648-010-9126-7>
- Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17(6), 722–738. <https://doi.org/10.1016/j.learninstruc.2007.09.013>
- Holmqvist, K., & Andersson, R. (2017). *Eye tracking: A comprehensive guide to methods, paradigms, and measures* (2nd edition). CreateSpace.
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. OUP Oxford.
- Hooge, I. T. C., Niehorster, D. C., Nyström, M., Andersson, R., & Hessels, R. S. (2022). Fixation classification: How to merge and select fixation candidates. *Behavior Research Methods*, 54(6), 2765–2776. <https://doi.org/10.3758/s13428-021-01723-1>
- Howe, F., & Knutzen, S. (2013). *Digitale Medien in der gewerblich-technischen Berufsausbildung. Einsatzmöglichkeiten digitaler Medien in Lern- und Arbeitsaufgaben [Digital media in industrial-technical vocational training. Possible uses of digital media in learning and work tasks]*. Bundesinstitut für Berufsbildung. https://www.bibb.de/dokumente/pdf/Expertise_Howe_Knutzen.pdf
- Hoyt, G. M., & McGoldrick, K. (2019). 50 years of economic instruction in the Journal of Economic Education. *The Journal of Economic Education*, 50(2), 168–195. <https://doi.org/10.1080/00220485.2019.1582388>

- Huang, Y., Miller, K. F., Cortina, K. S., & Richter, D. (2023). Teachers' professional vision in action: Comparing expert and novice teacher's real-life eye movements in the classroom. *Zeitschrift Für Pädagogische Psychologie*, 37(1–2), 122–139. <https://doi.org/10.1024/1010-0652/a000313>
- Huber, S. G., Günther, P. S., Schneider, N., Helm, C., Schwander, M., Schneider, J., & Pruitt, J. (2020). *COVID-19 und aktuelle Herausforderungen in Schule und Bildung [COVID-19 and current challenges in schools and education]*. Waxmann. <https://doi.org/10.31244/9783830942160>
- Hughes, J., Thomas, R., & Scharber, C. (2006). Assessing technology integration: The RAT–replacement, amplification, and transformation–framework. *Society for Information Technology & Teacher Education International Conference*, 1616–1620.
- Huß, B., & Dölle, F. (2021). *Bedeutung, Nutzung und Zugang zu Lehrbüchern an Hochschulen: Zentrale Ergebnisse der Befragungsstudie für das Bundesministerium für Bildung und Forschung (BMBF): Bericht Dezember 2021 [Importance, use and access to textbooks at universities : key findings of the survey study for the Federal Ministry of Education and Research (BMBF): Report December 2021]*. Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW). <https://doi.org/10.2314/KXP:1788361857>
- Huwer, J., Irion, T., Kuntze, S., Schaal, S., & Thyssen, C. (2019a). From TPaCK to DPaCK–Digitalization in education requires more than technical knowledge. In M. Shelley & S. A. Kiray (Eds.), *Education research highlights in mathematics, science and technology* (pp. 298–309). ISRES Publishing.
- Huwer, J., Irion, T., Kuntze, S., Schaal, S., & Thyssen, C. (2019b). Von TPaCK zu DPaCK–Digitalisierung im Unterricht erfordert mehr als technisches Wissen [From TPaCK to DPaCK digitization in the classroom requires more than technical knowledge]. *MNU Journal*, 72(5), 358–364.
- Hwang, G.-J., & Chang, C.-Y. (2021). A review of opportunities and challenges of chatbots in education. *Interactive Learning Environments*, 31(7), 4099–4112. <https://doi.org/10.1080/10494820.2021.1952615>
- Ifenthaler, D. (2014). AKOVIA: Automated Knowledge Visualization and Assessment. *Technology, Knowledge and Learning*, 19, 241–248. <https://doi.org/10.1007/s10758-014-9224-6>
- Ifenthaler, D. (2022). Automated Essay Scoring Systems. In O. Zawacki-Richter & I. Jung (Eds.), *Handbook of Open, Distance and Digital Education* (pp. 1–15). Springer. https://doi.org/10.1007/978-981-19-0351-9_59-1
- Ifenthaler, D., & Delcker, J. (2021). Check-up Distance Learning. Datengestützte Schul- und Unterrichtsentwicklung für berufliche Schulen vor Ort [Check-up Distance Learning. Data-supported school and teaching development for local vocational schools]. *Schulverwaltung BW*, 30(11), 296–299.
- Ifenthaler, D., Majumdar, R., Gorissen, P., Judge, M., Mishra, S., Raffaghelli, J., & Shimada, A. (2024). Artificial Intelligence in Education: Implications for Policymakers, Researchers, and Practitioners. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-024-09747-0>
- Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development*, 64(5), 923–938. <https://doi.org/10.1007/s11423-016-9477-y>

- Ifenthaler, D., & Schumacher, C. (2023). Reciprocal issues of artificial and human intelligence in education. *Journal of Research on Technology in Education*, 55(1), 1–6. <https://doi.org/10.1080/15391523.2022.2154511>
- Ifenthaler, D., & Tracey, M. W. (2016). Exploring the relationship of ethics and privacy in learning analytics and design: Implications for the field of educational technology. *Educational Technology Research and Development*, 64(5), 877–880. <https://doi.org/10.1007/s11423-016-9480-3>
- iMotions (9.3). (2023). [Computer software]. iMotions A/S
- Jarodzka, H. (2021). Research Methods in Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 41–54). Cambridge University Press. <https://doi.org/10.1017/9781108894333.006>
- Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in Educational Science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research*, 10(1), 1–18. <https://doi.org/10.16910/JEMR.10.1.3>
- Jaspersen, J., Carter, P., & Zmud, R. (2005). A Comprehensive Conceptualization of Post-Adoptive Behaviors Associated with Information Technology Enabled Work Systems. *MIS Quarterly*, 29(3), 525. <https://doi.org/10.2307/25148694>
- Jiang, J.-Y., Guo, F., Chen, J.-H., Tian, X.-H., & Lv, W. (2019). Applying eye-tracking technology to measure interactive experience toward the navigation interface of mobile games considering different visual attention mechanisms. *Applied Sciences*, 9(16), 3242.
- Jiang, Z., Xu, F. F., Araki, J., & Neubig, G. (2020). How Can We Know What Language Models Know? *Transactions of the Association for Computational Linguistics*, 8, 423–438.
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research and Development*, 48(4), 63–85. <https://doi.org/10.1007/BF02300500>
- Jonassen, D. H., & Hung, W. (2008). All Problems are Not Equal: Implications for Problem-Based Learning. *Interdisciplinary Journal of Problem-Based Learning*, 2(2), 6–28. <https://doi.org/10.7771/1541-5015.1080>
- Joshi, M., Choi, E., Weld, D. S., & Zettlemoyer, L. (2017). TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension. *arXiv:1705.03551 [Cs]*. <http://arxiv.org/abs/1705.03551>
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4), 329–354. <https://doi.org/10.1037/0033-295X.87.4.329>
- Kalyuga, S. (2013). Effects of Learner Prior Knowledge and Working Memory Limitations on Multimedia Learning. *Procedia - Social and Behavioral Sciences*, 83, 25–29. <https://doi.org/10.1016/j.sbspro.2013.06.005>
- Kalyuga, S., & Sweller, J. (2014). The Redundancy Principle in Multimedia Learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 247–262). Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369.013>
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling Laws for Neural Language Models. *arXiv:2001.08361 [Cs, Stat]*. <http://arxiv.org/abs/2001.08361>
- Kärner, T., Minkley, N., Rausch, A., Schley, T., & Sembill, D. (2018). Stress and Resources in Vocational Problem Solving. *Vocations and Learning*, 11(3), 365–398. <https://doi.org/10.1007/s12186-017-9193-8>

- Klein, P., Küchemann, S., Brückner, S., Zlatkin-Troitschanskaia, O., & Kuhn, J. (2019). Student Understanding of Graph Slope and Area under a Curve: A Replication Study Comparing First-Year Physics and Economics Students. *Physical Review Physics Education Research*, 15(2), 020116. <https://doi.org/10.1103/PhysRevPhysEducRes.15.020116>
- Kang, Z., & Landry, S. J. (2014). Using scanpaths as a learning method for a conflict detection task of multiple target tracking. *Human Factors*, 56(6), 1150–1162. <https://doi.org/10.1177/0018720814523066>
- Kärner, T., Minkley, N., Rausch, A., Schley, T., & Sembill, D. (2018). Stress and Resources in Vocational Problem Solving. *Vocations and Learning*, 11(3), 365–398. <https://doi.org/10.1007/s12186-017-9193-8>
- Ke, Z., & Ng, V. (2019). Automated Essay Scoring: A Survey of the State of the Art. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 6300–6308. <https://doi.org/10.24963/ijcai.2019/879>
- Kerres, M., & Kalz, M. (2003). Mediendidaktik in der Lehrerbildung [Media didactics in teacher training]. *Beiträge zur Lehrerinnen-und Lehrerbildung*, 21(3), 410–421.
- Keskin, Ö., Seidel, T., Stürmer, K., & Gegenfurtner, A. (2024). Eye-tracking research on teacher professional vision: A meta-analytic review. *Educational Research Review*, 42, 100586. <https://doi.org/10.1016/j.edurev.2023.100586>
- Ketamo, H. (2018). *Dreams and reality: How AI will change education*. UNESCO: United Nations Educational, Scientific and Cultural Organisation. <https://mgiep.unesco.org/article/dreams-and-reality-how-ai-will-change-education>
- Klatt, C., Härtel, S., & Markert, J. (2023). Digitalisierung im Berufsfeld Ernährung und Hauswirtschaft – Einsatz eines Property-Management-Systems aus der Hotellerie in der Lehramtsbildung für berufsbildende Schulen [Digitization in the Vocational Field of Nutrition and Home Economics—Implementation of a property management system adapted from the hotel industry in teacher education for vocational schools]. *bwp@ Berufs- und Wirtschaftspädagogik – online*, 45, 1–20.
- Knigge, M., Prifti, L., Kienegger, H., & Krcmar, H. (2017). Teaching enterprise organization and enterprise resource planning systems in schools: Playing a serious game with pupils. *2017 IEEE Global Engineering Education Conference (EDUCON)*, 486–495. <https://doi.org/10.1109/EDUCON.2017.7942891>
- Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60–70.
- Kohl, C., McIntosh, E. J., Unger, S., Haddaway, N. R., Kecke, S., Schiemann, J., & Wilhelm, R. (2018). Online tools supporting the conduct and reporting of systematic reviews and systematic maps: A case study on CADIMA and review of existing tools. *Environmental Evidence*, 7, 8. <https://doi.org/10.1186/s13750-018-0115-5>
- Kosel, C., Voggenreiter, A., Pfeffer, J., & Seidel, T. (2023). Measuring Teachers' Visual Expertise Using the Gaze Relational Index Based on Real-World Eyetracking Data and Varying Velocity Thresholds. *Journal of Expertise*, 6(2), 267–281.
- Kourilsky, M., & Wittrock, M. C. (1987). Verbal and graphical strategies in the teaching of economics. *Teaching and Teacher Education*, 3(1), 1–12. [https://doi.org/10.1016/0742-051X\(87\)90030-8](https://doi.org/10.1016/0742-051X(87)90030-8)
- Kramer, M., Förtsch, C., Aufleger, M., & Neuhaus, B. J. (2019). Der Einsatz digitaler Medien im gymnasialen Biologieunterricht: Eine deskriptive Auswertung einer quantitativen

- Videostudie [The use of digital media in secondary school biology lessons: A descriptive evaluation of a quantitative video study]. *Zeitschrift für Didaktik der Naturwissenschaften*, 25(1), 131–160. <https://doi.org/10.1007/s40573-019-00096-5>
- Krippendorff, K. (2004). Reliability in Content Analysis.: Some Common Misconceptions and Recommendations. *Human Communication Research*, 30(3), 411–433. <https://doi.org/10.1111/j.1468-2958.2004.tb00738.x>
- Kultusministerkonferenz. (2019a). *Rahmenlehrplan für den Ausbildungsberuf Kaufmann für Groß- und Außenhandelsmanagement und Kauffrau für Groß- und Außenhandelsmanagement [Framework curriculum for the training occupation of management assistant in wholesale and foreign trade management]*. https://www.kmk.org/themen/berufliche-schulen/duale-berufsausbildung/downloadbereich-rahmenlehrpläne.html?type=150&tx_fedownloads_pi1%5Bdownload%5D=45053&tx_fedownloads_pi1%5Baction%5D=forceDownload&tx_fedownloads_pi1%5Bcontroller%5D=Downloads&cHash=c5732d9a98f6f4837cf89b90b84ecc7f
- Kultusministerkonferenz. (2019b). *Rahmenlehrplan für die Ausbildungsberufe Kaufmann für IT-System-Management und Kauffrau für IT-System-Management Kaufmann für Digitalisierungsmanagement und Kauffrau für Digitalisierungsmanagement [Framework curriculum for the training occupations of IT systems management clerk and IT systems management clerk Digitization management clerk]*. https://www.kmk.org/themen/berufliche-schulen/duale-berufsausbildung/downloadbereich-rahmenlehrpläne.html?type=150&tx_fedownloads_pi1%5Bdownload%5D=45053&tx_fedownloads_pi1%5Baction%5D=forceDownload&tx_fedownloads_pi1%5Bcontroller%5D=Downloads&cHash=c5732d9a98f6f4837cf89b90b84ecc7f
- Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Lee, M.-H., Chiou, G.-L., Liang, J.-C., & Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review*, 10, 90–115. <https://doi.org/10.1016/j.edurev.2013.10.001>
- Lajoie, S. P., & Naismith, L. (2012). Computer-Based Learning Environments. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 716–718). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_512
- Lameras, P., & Arnab, S. (2021). Power to the Teachers: An Exploratory Review on Artificial Intelligence in Education. *Information*, 13(1), 14. <https://doi.org/10.3390/info13010014>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Lassak, M. (2009). Using dynamic graphs to reveal student reasoning. *International Journal of Mathematical Education in Science and Technology*, 40(5), 690–696. <https://doi.org/10.1080/00207390902759642>
- Leahy, C. (2022, February 2). *Announcing GPT-NeoX-20B*. EleutherAI Blog. <https://blog.eleuther.ai/announcing-20b/>
- Lee, J. Y., Donkers, J., Jarodzka, H., & Merriënboer, J. van. (2019). How prior knowledge affects problem-solving performance in a medical simulation game: Using game-logs and eye-tracking. *Computers in Human Behavior*, 99, 268–277. <https://doi.org/10.1016/j.chb.2019.05.035>
- Lee, J. Y., Donkers, J., Jarodzka, H., Sellenraad, G., & van Merriënboer, J. J. G. (2020). Different effects of pausing on cognitive load in a medical simulation game. *Computers in Human Behavior*, 110, 106385. <https://doi.org/10.1016/j.chb.2020.106385>

- Legner, C., Eymann, T., Hess, T., Matt, C., Böhmman, T., Drews, P., Mädche, A., Urbach, N., & Ahlemann, F. (2017). Digitalization: Opportunity and Challenge for the Business and Information Systems Engineering Community. *Business & Information Systems Engineering*, 59(4), 301–308. <https://doi.org/10.1007/s12599-017-0484-2>
- Leiß, T. V., Rausch, A., & Seifried, J. (2022). Problem-Solving and Tool Use in Office Work: The Potential of Electronic Performance Support Systems to Promote Employee Performance and Learning. *Frontiers in Psychology*, 13, 869428. <https://doi.org/10.3389/fpsyg.2022.869428>
- Lester, B., Al-Rfou, R., & Constant, N. (2021). The Power of Scale for Parameter-Efficient Prompt Tuning. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 3045–3059.
- Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv Preprint arXiv:1910.13461*.
- Liang, G., On, B.-W., Jeong, D., Kim, H.-C., & Choi, G. (2018). Automated Essay Scoring: A Siamese Bidirectional LSTM Neural Network Architecture. *Symmetry*, 10(12), 682. <https://doi.org/10.3390/sym10120682>
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ (Clinical Research Ed.)*, 339, b2700. <https://doi.org/10.1136/bmj.b2700>
- Lind Pantzare, A. (2015). Interrater reliability in large-scale assessments—Can teachers score national tests reliably without external controls? *Practical Assessment, Research, and Evaluation*, 20(1), 9.
- Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., & Neubig, G. (2021). Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *arXiv:2107.13586 [Cs]*. <http://arxiv.org/abs/2107.13586>
- Liu, X., Ji, K., Fu, Y., Du, Z., Yang, Z., & Tang, J. (2021). P-Tuning v2: Prompt Tuning Can Be Comparable to Fine-tuning Universally Across Scales and Tasks. *arXiv Preprint arXiv:2110.07602*.
- Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *arXiv:1907.11692 [Cs]*. <http://arxiv.org/abs/1907.11692>
- Lubega, J., Kajura, M., & Muyinda, P. B. (2014). Adoption of the SAMR Model to Assess ICT Pedagogical Adoption: A Case of Makerere University. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 4(2), 106–115. <https://doi.org/10.7763/IJEEEE.2014.V4.312>
- Lucy, L., & Bamman, D. (2021). Gender and Representation Bias in GPT-3 Generated Stories. *Proceedings of the Third Workshop on Narrative Understanding*, 48–55. <https://doi.org/10.18653/v1/2021.nuse-1.5>
- Ludwig, S., Mayer, C., Hansen, C., Eilers, K., & Brandt, S. (2021). Automated essay scoring using transformer models. *Psych*, 3(4), 897–915.
- Ludwig, S., Rausch, A., Deutscher, V., & Seifried, J. (2024). Predicting problem-solving success in an office simulation applying N-grams and a random forest to behavioral process data. *Computers & Education*, 218, 105093. <https://doi.org/10.1016/j.compedu.2024.105093>

- Maddox, B., Bayliss, A. P., Fleming, P., Engelhardt, P. E., Edwards, S. G., & Borgonovi, F. (2018). Observing response processes with eye tracking in international large-scale assessments: Evidence from the OECD PIAAC assessment. *European Journal of Psychology of Education, 33*(3), 543–558. <https://doi.org/10.1007/s10212-018-0380-2>
- Mao, J. (2017). New Technologies and Old Professional Development: A Reflection on Emerging Approaches. In W. W. K. Ma, C.-K. Chan, K. Tong, H. Fung, & C. W. R. Fong (Eds.), *New Ecology for Education—Communication X Learning* (pp. 67–77). Springer. https://doi.org/10.1007/978-981-10-4346-8_6
- Marín, V. I., & Castañeda, L. (2023). Developing Digital Literacy for Teaching and Learning. In O. Zawacki-Richter & I. Jung (Eds.), *Handbook of Open, Distance and Digital Education* (pp. 1089–1108). Springer. https://doi.org/10.1007/978-981-19-2080-6_64
- Marks-Anglin, A., & Chen, Y. (2020). A historical review of publication bias. *Research Synthesis Methods, 11*(6), 725–742. <https://doi.org/10.1002/jrsm.1452>
- Mayer, C. W. (2022). Lehren und Lernen mit Enterprise Resource Planning (ERP) Systemen—Typische Bearbeitungsprobleme als Grundlage der Lernprozessgestaltung [Teaching and learning with Enterprise Resource Planning (ERP) systems Typical processing problems as a basis for designing the learning process]. *Bwp@ Berufs- Und Wirtschaftspädagogik – Online, 43*, 1–17.
- Mayer, C. W., Findeisen, S., Guggemos, J., & Seifried, J. (under review). Visual attention while solving economic learning tasks using dynamic graphs – An eye-tracking study. *Journal of Economic Education*.
- Mayer, C. W., Gentner, S., & Seifried, J. (2023). Digitale Unterrichtspraxis an kaufmännischen Schulen in der Corona-Pandemie—Eine Momentaufnahme [Digital teaching practice at commercial schools during the corona pandemic - a snapshot]. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 119*(3), 531–552. <https://doi.org/10.25162/zbw-2023-0021>
- Mayer, C. W., Ludwig, S., & Brandt, S. (2022). Prompt text classifications with transformer models! An exemplary introduction to prompt-based learning with large language models. *Journal of Research on Technology in Education, 55*(1), 1–17. <https://doi.org/10.1080/15391523.2022.2142872>
- Mayer, C. W., Rausch, A., & Seifried, J. (2023). Analysing domain-specific problem-solving processes within authentic computer-based learning and training environments by using eye-tracking: A scoping review. *Empirical Research in Vocational Education and Training, 15*(1), 2. <https://doi.org/10.1186/s40461-023-00140-2>
- Mayer, C. W., & Seifried, J. (2022). Lehren und Lernen mit ERP-Systemen – Befunde der Evaluation einer Lehrveranstaltung zur Förderung der Technologiekompetenz von Studierenden der Wirtschaftspädagogik [Teaching and learning with ERP systems—Findings from the evaluation of a course to promote the technological competence of business education students]. In U. Mathis, N. Ondrusch, D. Kilian, H. Krcmar, K. Turowski, & S. Weidner (Eds.), *Proceedings of the SAP Academic Community Conference 2022 DACH* (pp. 10–24). Technische Universität München. <https://mediatum.ub.tum.de/doc/1685828/document.pdf#page=10>
- Mayer, C. W., & Seifried, J. (2024). Looking behind the scenes—Being able to master ERP systems as a goal of vocational education and training. In Cedefop & OECD (Eds.), *Apprenticeships and the digital transition* (pp. 118–127). Publications Office.
- Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional Science, 26*(1/2), 49–63. <https://doi.org/10.1023/A:1003088013286>

- Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction, 20*(2), 167–171. <https://doi.org/10.1016/j.learninstruc.2009.02.012>
- Mayer, R. E. (2021). *Multimedia Learning* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781316941355>
- Mayer, R. E., & Wittrock, M. C. (2006). Problem Solving. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of Educational Psychology* (2nd ed., pp. 287–303). Routledge.
- McPherson, M. S., & Bacow, L. S. (2015). Online Higher Education: Beyond the Hype Cycle. *Journal of Economic Perspectives, 29*(4), 135–154. <https://doi.org/10.1257/jep.29.4.135>
- Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. U.S. Department of Education. www.ed.gov/about/offices/list/opepd/ppss/reports.html
- Metcalf, J. (2017). Learning from Errors. *Annual Review of Psychology, 68*(1), 465–489. <https://doi.org/10.1146/annurev-psych-010416-044022>
- Mishra, P. (2019). Considering Contextual Knowledge: The TPACK Diagram Gets an Upgrade. *Journal of Digital Learning in Teacher Education, 35*(2), 76–78. <https://doi.org/10.1080/21532974.2019.1588611>
- Mishra, P., & Koehler, M. J. (2006). Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge. *Teachers College Record: The Voice of Scholarship in Education, 108*(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Models—Hugging Face*. (n.d.). Retrieved March 30, 2022, from <https://huggingface.co/models>
- Molenaar, I. (2022). The concept of hybrid human-AI regulation: Exemplifying how to support young learners' self-regulated learning. *Computers and Education: Artificial Intelligence, 3*, 100070. <https://doi.org/10.1016/j.caeai.2022.100070>
- Molenda, M. H. (2022). History and Development of Instructional Design and Technology. In O. Zawacki-Richter & J. Insung (Eds.), *Handbook of Open, Distance and Digital Education* (pp. 1–18). Springer. https://doi.org/10.1007/978-981-19-0351-9_4-1
- Møller, F., Laursen, M. L., & Sjølie, A. K. (2006). The contribution of microsaccades and drifts in the maintenance of binocular steady fixation. *Graefes Archive for Clinical and Experimental Ophthalmology, 244*(4), 465–471. <https://doi.org/10.1007/s00417-004-0989-5>
- Mollick, E. (2024). *Co-intelligence: Living and working with AI*. W H Allen.
- Moore, M. G., & Kearsley, G. G. (1996). *Distance education: A system view*. Wadsworth.
- Mouza, C., Karchmer-Klein, R., Nandakumar, R., Yilmaz Ozden, S., & Hu, L. (2014). Investigating the impact of an integrated approach to the development of preservice teachers' technological pedagogical content knowledge (TPACK). *Computers & Education, 71*, 206–221. <https://doi.org/10.1016/j.compedu.2013.09.020>
- Mugge, P., Abbu, H., Michaelis, T. L., Kwiatkowski, A., & Gudergan, G. (2020). Patterns of Digitization: A Practical Guide to Digital Transformation. *Research-Technology Management, 63*(2), 27–35. <https://doi.org/10.1080/08956308.2020.1707003>
- Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a

- systematic or scoping review approach. *BMC Medical Research Methodology*, *18*(1), 143. <https://doi.org/10.1186/s12874-018-0611-x>
- Munshi, A., Mishra, S., Zhang, N., Paquette, L., Ocumpaugh, J., Baker, R., & Biswas, G. (2020). Modeling the Relationships Between Basic and Achievement Emotions in Computer-Based Learning Environments. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial Intelligence in Education* (Vol. 12163, pp. 411–422). Springer International Publishing. https://doi.org/10.1007/978-3-030-52237-7_33
- Nakayama, M., & Santiago, R. (2012). Learner Characteristics and Online Learning. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 1745–1747). Springer. https://doi.org/10.1007/978-1-4419-1428-6_1858
- Naqvi, S., Raza, M., Ghazal, S., Salehi, S., Kang, Z., & Teodoriu, C. (2020). Simulation-based training to enhance process safety in offshore energy operations: Process tracing through eye-tracking. *Process Safety and Environmental Protection*, *138*, 220–235. <https://doi.org/10.1016/j.psep.2020.03.016>
- Ndukwe, I. G., Daniel, B. K., & Amadi, C. E. (2019). A Machine Learning Grading System Using Chatbots. In S. Isotani, E. Millán, A. Ogan, P. Hastings, B. McLaren, & R. Luckin (Eds.), *Artificial Intelligence in Education* (pp. 365–368). Springer. https://doi.org/10.1007/978-3-030-23207-8_67
- Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, *15*(3), 263–280. <https://doi.org/10.1002/ejsp.2420150303>
- Negi, S., & Mitra, R. (2020). Fixation duration and the learning process: An eye tracking study with subtitled videos. *Journal of Eye Movement Research*, *13*(6), 1–15. <https://doi.org/10.16910/jemr.13.6.1>
- Niegemann, H. (2020). Introduction—A Look Back Ahead. In E. Wuttke, J. Seifried, & H. Niegemann (Eds.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (pp. 15–19). Barbara Budrich. <https://doi.org/10.2307/j.ctv18dvv1c>
- Niegemann, H., & Weinberger, A. (2020). Was ist Bildungstechnologie? [What is educational technology?]. In H. Niegemann & A. Weinberger (Eds.), *Handbuch Bildungstechnologie* (pp. 3–16). Springer. https://doi.org/10.1007/978-3-662-54368-9_1
- OECD & The World Bank. (2022). *How Learning Continued during the COVID-19 Pandemic: Global Lessons from Initiatives to Support Learners and Teachers* (S. Vincent-Lancrin, C. Cobo Romani, & F. Reimers, Eds.). OECD Publishing. <https://doi.org/10.1787/bbeca162-en>
- Olsen, A., & Matos, R. (2012). Identifying parameter values for an I-VT fixation filter suitable for handling data sampled with various sampling frequencies. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 317–320). Association for Computing Machinery. <https://doi.org/10.1145/2168556.2168625>
- O'Meara, P., Munro, G., Williams, B., Cooper, S., Bogossian, F., Ross, L., Sparkes, L., Browning, M., & McClounan, M. (2015). Developing situation awareness amongst nursing and paramedicine students utilizing eye tracking technology and video debriefing techniques: A proof of concept paper. *International Emergency Nursing*, *23*(2), 94–99. <https://doi.org/10.1016/j.ienj.2014.11.001>
- Orquin, J. L., & Mueller Loose, S. (2013). Attention and choice: A review on eye movements in decision making. *Acta Psychologica*, *144*(1), 190–206. <https://doi.org/10.1016/j.actpsy.2013.06.003>

- Orwin, R. G., & Vevea, J. L. (2009). Evaluating coding decisions. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 177–203). Russell Sage Foundation.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. <https://doi.org/10.1136/bmj.n71>
- Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford university press.
- Paperno, D., Kruszewski, G., Lazaridou, A., Pham, N.-Q., Bernardi, R., Pezzelle, S., Baroni, M., Boleda, G., & Fernández, R. (2016). The LAMBADA dataset: Word prediction requiring a broad discourse context. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, 1525–1534.
- Papers with Code—The latest in Machine Learning*. (n.d.). Retrieved March 21, 2022, from <https://paperswithcode.com/>
- Parviainen, P., Tihinen, M., Kääriäinen, J., & Teppola, S. (2017). Tackling the digitalization challenge: How to benefit from digitalization in practice. *International Journal of Information Systems and Project Management*, *5*(1), 63–77. <https://doi.org/10.12821/ijispm050104>
- Peebles, D., & Cheng, P. C.-H. (2003). Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors*, *45*(1), 28–46.
- Pellegrino, J. W. (2002). Knowing what students know. *Issues in Science and Technology*, *19*(2), 48–52.
- Pellegrino, J. W. (2006). Rethinking and Redesigning Curriculum, Instruction and Assessment: What Contemporary Research and Theory Suggests. *Commission on the Skills of the American Workforce*, Chicago, 1–15.
- Penso, S. (2002). Pedagogical Content Knowledge: How do student teachers identify and describe the causes of their pupils' learning difficulties? *Asia-Pacific Journal of Teacher Education*, *30*(1), 25–37. <https://doi.org/10.1080/13598660120114959>
- Peterson, R. A., & Cavanaugh, J. E. (2020). Ordered quantile normalization: A semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics*, *47*(13–15), 2312–2327. <https://doi.org/10.1080/02664763.2019.1630372>
- Pethig, F., & Kroenung, J. (2022). Biased Humans, (Un)Biased Algorithms? *Journal of Business Ethics*, *183*, 637–652. <https://doi.org/10.1007/s10551-022-05071-8>
- Petko, D. (2012). Teachers' pedagogical beliefs and their use of digital media in classrooms: Sharpening the focus of the 'will, skill, tool' model and integrating teachers' constructivist orientations. *Computers & Education*, *58*(4), 1351–1359. <https://doi.org/10.1016/j.compedu.2011.12.013>
- Petroni, F., Rocktäschel, T., Riedel, S., Lewis, P., Bakhtin, A., Wu, Y., & Miller, A. (2019). Language Models as Knowledge Bases? *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2463–2473.
- Pinkwart, N. (2016). Another 25 Years of AIED? Challenges and Opportunities for Intelligent Educational Technologies of the Future. *International Journal of Artificial Intelligence in Education*, *26*(2), 771–783. <https://doi.org/10.1007/s40593-016-0099-7>

- Ploetzner, R., & Lowe, R. (2004). Dynamic visualisations and learning. *Learning and Instruction, 14*(3), 235–240. <https://doi.org/10.1016/j.learninstruc.2004.06.001>
- Pongratz, H., Tramm, P. T., & Wilbers, K. (Eds.). (2009). *Prozessorientierte Wirtschaftsdidaktik und Einsatz von ERP-Systemen im kaufmännischen Unterricht [Process-oriented business didactics and the use of ERP systems in commercial teaching]*. Shaker.
- Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning, 12*, 22. <https://doi.org/10.1186/s41039-017-0062-8>
- Posit Team. (2023). *RStudio: Integrated Development Environment for R* [Computer software]. Posit Software, PBC. <http://www.posit.co/>
- Premsky, M. (2001a). Digital Natives, Digital Immigrants Part 1. *On the Horizon, 9*(5), 1–6. <https://doi.org/10.1108/10748120110424816>
- Premsky, M. (2001b). Digital Natives, Digital Immigrants Part 2: Do They Really Think Differently? *On the Horizon, 9*(6), 1–6. <https://doi.org/10.1108/10748120110424843>
- Psotka, J. (2012). Interactive Learning Environments. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 1604–1606). Springer. https://doi.org/10.1007/978-1-4419-1428-6_321
- Puentedura, R. (2006). *Transformation, Technology, and Education*. <http://hippasus.com/resources/tte/>
- Puentedura, R. (2014). *SAMR, Learning, and Assessment*. <http://www.hippasus.com/rrpweblog/archives/2014/11/28/SAMRLearningAssessment.pdf>
- Puri, R., & Catanzaro, B. (2019). Zero-shot Text Classification With Generative Language Models. *arXiv:1912.10165 [Cs]*. <http://arxiv.org/abs/1912.10165>
- Pushp, P. K., & Srivastava, M. M. (2017). Train once, test anywhere: Zero-shot learning for text classification. *arXiv Preprint arXiv:1712.05972*.
- Puustinen, M., Sääntti, J., Koski, A., & Tammi, T. (2018). Teaching: A practical or research-based profession? Teacher candidates' approaches to research-based teacher education. *Teaching and Teacher Education, 74*, 170–179. <https://doi.org/10.1016/j.tate.2018.05.004>
- Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained models for natural language processing: A survey. *Science China Technological Sciences, 63*(10), 1872–1897. <https://doi.org/10.1007/s11431-020-1647-3>
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language, 59*(4), 413–425. <https://doi.org/10.1016/j.jml.2008.02.002>
- R Core Team. (2021). *R: A language and environment for statistical computing*. [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>
- R Studio Team. (2021). *RStudio: Integrated Development Environment for R*. [Computer software]. RStudio, PBC. <http://www.rstudio.com/>
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving language understanding by generative pre-training. *OpenAI Blog*.
- Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*.

- Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. J. (2019). *Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer*. <https://doi.org/10.48550/ARXIV.1910.10683>
- Ramineni, C., Trapani, C. S., Williamson, D. M., Davey, T., & Bridgeman, B. (2012). Evaluation of the e-rate® scoring engine for the TOEFL® independent and integrated prompts. *ETS Research Report Series*, 2012(1), i–51. <https://doi.org/10.1002/j.2333-8504.2012.tb02288.x>
- Rausch, A., Deutscher, V., Seifried, J., Brandt, S., & Winther, E. (2021). Die web-basierte Bürosimulation LUCA – Funktionen, Einsatzmöglichkeiten und Forschungsausblick [The web-based office simulation LUCA - functions, possible applications and research outlook]. *Zeitschrift für Berufsund Wirtschaftspädagogik*, 117(3), 372–394. <https://doi.org/10.25162/zbw-2021-0017>
- Rausch, A., Kögler, K., & Seifried, J. (2019). Validation of Embedded Experience Sampling (EES) for Measuring Non-cognitive Facets of Problem-Solving Competence in Scenario-Based Assessments. *Frontiers in Psychology*, 10, 1200. <https://doi.org/10.3389/fpsyg.2019.01200>
- Rausch, A., Seifried, J., Wuttke, E., Kögler, K., & Brandt, S. (2016). Reliability and validity of a computer-based assessment of cognitive and non-cognitive facets of problem-solving competence in the business domain. *Empirical Research in Vocational Education and Training*, 8(1), 9. <https://doi.org/10.1186/s40461-016-0035-y>
- Rausch, A., & Wuttke, E. (2016). Development of a Multi-Faceted Model of Domain-Specific Problem-Solving Competence and its Acceptance by Different Stakeholders in the Business Domain. *Unterrichtswissenschaft*, 44(2), 169–184.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.
- Rayner, K., & Castelano, M. (2007). Eye Movements during Reading, Scene Perception, Visual Search, and While Looking at Print Advertisements. In M. Wedel & R. Pieters (Eds.), *Visual Marketing*. Psychology Press. <https://doi.org/10.4324/9780203809617>
- Redecker, C. (2017). *European framework for the digital competence of educators: DigCompEdu*. (Y. Punie, Ed.). Publications Office of the European Union. <https://data.europa.eu/doi/10.2760/178382>
- Reich, J. (2020). *Failure to disrupt: Why technology alone can't transform education*. Harvard University Press.
- Richter, J., Wehrle, A., & Scheiter, K. (2021). How the poor get richer: Signaling guides attention and fosters learning from text-graph combinations for students with low, but not high prior knowledge. *Applied Cognitive Psychology*, 35(3), 632–645.
- Rieber, L. P. (1991). Animation, incidental learning, and continuing motivation. *Journal of Educational Psychology*, 83(3), 318–328. <https://doi.org/10.1037/0022-0663.83.3.318>
- Rodgers, T. L., Bolton, P., & Van Der Gryp, P. (2023). An interactive graph resource for chemical engineering teaching. *Education for Chemical Engineers*, 45, 80–89. <https://doi.org/10.1016/j.ece.2023.08.004>
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33, 1–39. <https://doi.org/10.1007/s10462-009-9124-7>
- Roll, M., & Ifenthaler, D. (2020). The Impact of Learning Factories on Multidisciplinary Digital Competencies. In E. Wuttke, J. Seifried, & H. Niegemann (Eds.), *Vocational Education*

- and Training in the Age of Digitization: Challenges and Opportunities* (pp. 23–38). Barbara Budrich. <https://doi.org/10.2307/j.ctv18dvv1c>
- Roll, M., & Ifenthaler, D. (2021). Learning Factories 4.0 in technical vocational schools: Can they foster competence development? *Empirical Research in Vocational Education and Training*, 13(1), 20. <https://doi.org/10.1186/s40461-021-00124-0>
- Romera-Paredes, B., & Torr, P. (2015). An embarrassingly simple approach to zero-shot learning. *Proceedings of the 32nd International Conference on Machine Learning*, 2152–2161. <https://proceedings.mlr.press/v37/romera-paredes15.html>
- Rosenberg, J. M., & Koehler, M. J. (2015). Context and Technological Pedagogical Content Knowledge (TPACK): A Systematic Review. *Journal of Research on Technology in Education*, 47(3), 186–210. <https://doi.org/10.1080/15391523.2015.1052663>
- Rosengrant, D., Herrington, D., & O'Brien, J. (2021). Investigating Student Sustained Attention in a Guided Inquiry Lecture Course Using an Eye Tracker. *Educational Psychology Review*, 33(1), 11–26. <https://doi.org/10.1007/s10648-020-09540-2>
- Ross, J. A., & Bruce, C. D. (2007). Teacher self-assessment: A mechanism for facilitating professional growth. *Teaching and Teacher Education*, 23(2), 146–159. <https://doi.org/10.1016/j.tate.2006.04.035>
- Ruf, V., Horrer, A., Berndt, M., Hofer, S. I., Fischer, F., Fischer, M. R., Zottmann, J. M., Kuhn, J., & Küchemann, S. (2023). A Literature Review Comparing Experts' and Non-Experts' Visual Processing of Graphs during Problem-Solving and Learning. *Education Sciences*, 13(2), 216. <https://doi.org/10.3390/educsci13020216>
- SAP4school IUS. (2023). <https://s4s.integrus.de/>
- Sainz, O., & Rigau, G. (2021). Ask2Transformers: Zero-Shot Domain labelling with Pretrained Language Models. *Proceedings of the 11th Global Wordnet Conference*, 44–52.
- Sänger, N., & Jenert, T. (2023). TPACK, DPACK, XY-PACK?: Eine kritische Auseinandersetzung mit der Modellierung digitaler Inhaltskompetenzen für die berufliche Bildung [TPACK, DPACK, XY-PACK?: A Critical Examination of the Modeling of Digital Content Competencies for Vocational Education]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 119(3), 434–454. <https://doi.org/10.25162/zbw-2023-0017>
- Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv Preprint arXiv:1910.01108*.
- Says, U. (2021, October 11). *Using DeepSpeed and Megatron to Train Megatron-Turing NLG 530B, the World's Largest and Most Powerful Generative Language Model*. NVIDIA Technical Blog. <https://developer.nvidia.com/blog/using-deepspeed-and-megatron-to-train-megatron-turing-nlg-530b-the-worlds-largest-and-most-powerful-generative-language-model/>
- Scao, T. L., & Rush, A. (2021). How many data points is a prompt worth? *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2627–2636. <https://doi.org/10.18653/v1/2021.naacl-main.208>
- Schäfer, P., Link, N., & Walker, F. (2020). Assessing Professional Knowledge of Teachers at Vocational Schools—Using the Example of a Professional Development for Automation and Digitized Production. In E. Wuttke, J. Seifried, & H. Niegemann (Eds.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (pp. 131–154). Barbara Budrich. <https://doi.org/10.2307/j.ctv18dvv1c>

- Scheiter, K. (2021). Lernen und Lehren mit digitalen Medien: Eine Standortbestimmung [Learning and teaching with digital media: An assessment of the current situation]. *Zeitschrift für Erziehungswissenschaft*, 24(5), 1039–1060. <https://doi.org/10.1007/s11618-021-01047-y>
- Scheiter, K., & Eitel, A. (2017). The Use of Eye Tracking as a Research and Instructional Tool in Multimedia Learning. In C. Was, F. Sansosti, & B. Morris (Eds.), *Eye-Tracking Technology Applications in Educational Research* (pp. 143–164). IGI Global. <https://doi.org/10.4018/978-1-5225-1005-5>
- Scherer, R., Tondeur, J., Siddiq, F., & Baran, E. (2018). The importance of attitudes toward technology for pre-service teachers' technological, pedagogical, and content knowledge: Comparing structural equation modeling approaches. *Computers in Human Behavior*, 80, 67–80. <https://doi.org/10.1016/j.chb.2017.11.003>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online*, 8(2), 23–74.
- Schick, T., & Schütze, H. (2021a). Exploiting Cloze-Questions for Few-Shot Text Classification and Natural Language Inference. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 255–269.
- Schick, T., & Schütze, H. (2021b). It's Not Just Size That Matters: Small Language Models Are Also Few-Shot Learners. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2339–2352.
- Schmid, M., Brianza, E., Mok, S. Y., & Petko, D. (2024). Running in circles: A systematic review of reviews on technological pedagogical content knowledge (TPACK). *Computers & Education*, 214, 105024. <https://doi.org/10.1016/j.compedu.2024.105024>
- Schmid, R. F., Bernard, R. M., Borokhovski, E., Tamim, R., Abrami, P. C., Wade, C. A., Surkes, M. A., & Lowerison, G. (2009). Technology's effect on achievement in higher education: A Stage I meta-analysis of classroom applications. *Journal of Computing in Higher Education*, 21(2), 95–109. <https://doi.org/10.1007/s12528-009-9021-8>
- Schmid, M., Brianza, E., & Petko, D. (2020). Developing a short assessment instrument for Technological Pedagogical Content Knowledge (TPACK.xs) and comparing the factor structure of an integrative and a transformative model. *Computers & Education*, 157, 103967. <https://doi.org/10.1016/j.compedu.2020.103967>
- Schmid, M., Krannich, M., & Petko, D. (2020). Technological Pedagogical Content Knowledge. Entwicklungen und Implikationen [Technological Pedagogical Content Knowledge. Developments and implications]. *Journal für LehrerInnenbildung*, 20(1), 116–124. <https://doi.org/10.25656/01:19396>
- Schneider, J., Schenk, B., & Niklaus, C. (2023). *Towards LLM-based Autograding for Short Textual Answers*. <https://doi.org/10.48550/ARXIV.2309.11508>
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. [https://doi.org/10.1016/S0959-4752\(02\)00017-8](https://doi.org/10.1016/S0959-4752(02)00017-8)
- Schrenk, R. (2021). Auswirkungen der COVID-19 Pandemie auf die Arbeit mit Moodle – Aktuelles aus Österreichs Schulen [Impact of the COVID-19 pandemic on working with Moodle - News from Austrian schools]. *GW-Unterricht*, 1, 54–60. <https://doi.org/10.1553/gw-unterricht162s54>

- Schrivers, A. T., Morrow, D. G., Wickens, C. D., & Talleur, D. A. (2008). Expertise differences in attentional strategies related to pilot decision making. *Human Factors, 50*(6), 864–878. <https://doi.org/10.1518/001872008X374974>
- Seel, N. M. (2012). Attitudes – Formation and Change. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 371–374). Springer. https://doi.org/10.1007/978-1-4419-1428-6_359
- Seifried, J. (2008). Methodische Gestaltung des Unterrichts an kaufmännischen Schulen [Methodical design of teaching at commercial schools]. *Wirtschaft & Erziehung, 60*(11), 364–370.
- Seifried, J. (2021). Der lange Weg in ein „New Normal“ – Noch ist unklar, was nach der Pandemie bleibt [The long road to a "new normal" - It is still unclear what will remain after the pandemic]. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 117*(3), 309. <https://doi.org/10.25162/zbw-2021-0014>
- Seifried, J., Beck, K., Ertelt, B.-J., & Frey, A. (Eds.). (2019). *Beruf, Beruflichkeit, Employability [occupation, professionalism, employability]*. wbv.
- Seifried, J., Brandt, S., Kögler, K., & Rausch, A. (2020). The computer-based assessment of domain-specific problem-solving competence—A three-step scoring procedure. *Cogent Education, 7*(1), 1719571. <https://doi.org/10.1080/2331186X.2020.1719571>
- Sembill, D., Rausch, A., Wuttke, E., Seifried, J., Wolf, K. D., Martens, T., & Brandt, S. (2016). *Domain specific Problem-solving Skills of Prospective Industrial Clerks (DomPL-IK) Modellierung und Messung domänenspezifischer Problemlösekompetenz bei Industriekaufleuten (DomPL-IK) (Version 1) [Dataset]*. IQB - Institute for Educational Quality Improvement. https://doi.org/10.5159/IQB_DOMPL-IK_V1
- Seufert, S., Guggemos, J., & Sailer, M. (2021). Technology-related knowledge, skills, and attitudes of pre- and in-service teachers: The current situation and emerging trends. *Computers in Human Behavior, 115*, 106552. <https://doi.org/10.1016/j.chb.2020.106552>
- Seufert, S., Guggemos, J., & Tarantini, E. (2018). Digitale Transformation in Schulen – Kompetenzanforderungen an Lehrpersonen [Digital transformation in schools - competence requirements for teachers]. *Beiträge zur Lehrerinnen- und Lehrerbildung, 36*(2), 175–193. <https://doi.org/10.25656/01:17096>
- Seufert, S., Guggemos, J., Tarantini, E., & Schumann, S. (2019). Professionelle Kompetenzen von Lehrpersonen im Kontext des digitalen Wandels: Entwicklung eines Rahmenkonzepts und Validierung in der kaufmännischen Domäne [Professional competencies of teachers in the context of digital change: development of a framework concept and validation in the commercial domain]. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 115*(2), 312–339. <https://doi.org/10.25162/zbw-2019-0013>
- Shermis, M. D., & Burstein, J. (Eds.). (2013). *Handbook on automated essay evaluation: Current applications and new directions*. Routledge.
- Shin, T., Razeghi, Y., Logan IV, R. L., Wallace, E., & Singh, S. (2020). AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 4222–4235.
- Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher, 15*(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Sierra, J. (2020). The importance of simulation in teaching and learning economics: The students' perspective. *Innovations in Education and Teaching International, 57*(5), 521–531. <https://doi.org/10.1080/14703297.2019.1647268>

- Silva, A., Tambwekar, P., & Gombolay, M. (2021). Towards a Comprehensive Understanding and Accurate Evaluation of Societal Biases in Pre-Trained Transformers. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2383–2389. <https://doi.org/10.18653/v1/2021.naacl-main.189>
- Sloane, P. F. E. (2020). Kaufmännische Berufsbildung nach Corona: Überlegungen für die zukünftige Lehrerbildung an beruflichen Schulen und Berufskollegs [Commercial vocational training after Corona: considerations for future teaching at vocational schools and vocational colleges]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 116(4), 533–543. <https://doi.org/10.25162/zbw-2020-0023>
- Smith, E. E., Kahlke, R., & Judd, T. (2020). Not just digital natives: Integrating technologies in professional education contexts. *Australasian Journal of Educational Technology*, 36(3), 1–14. <https://doi.org/10.14742/ajet.5689>
- Smith, F. J. (1913). The Evolution of the Motion Picture: VI—Looking into the Future with Thomas A. Edison. In *The New York Dramatic Mirror* (pp. 24, Column 3).
- Snell, J., Swersky, K., & Zemel, R. (2017). Prototypical networks for few-shot learning. *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, 4080–4090.
- Sohn, M., Douglass, S., Chen, M., & Anderson, J. (2005). Characteristics of fluent skills in a complex, dynamic problem-solving task. *Human Factors*, 47(4), 742–752. <https://doi.org/10.1518/001872005775570943>
- Sorden, S. D. (2013). The cognitive theory of multimedia learning. In B. J. Irby, G. Brown, R. Lara-Alecio, & S. Jackson (Eds.), *The handbook of educational theories*. (pp. 155–167). IAP.
- Spener, C., Häuber, G., Horlacher, T., & Schumann, S. (2019). Der berufsschulische Einsatz von ERP-Systemen in der Ausbildung von Industriekaufleuten [The vocational use of ERP systems in the training of industrial clerks]. In K. Wilbers (Ed.), *Digitale Transformation kaufmännischer Bildung: Ausbildung in Industrie und Handel hinterfragt* (pp. 195–216). epubli. https://www.pedocs.de/frontdoor.php?source_opus=17968
- Spener, C., & Schumann, S. (2021). Wissensseffekte des ERP-Einsatzes in der kaufmännischen Berufsschule [Knowledge effects of ERP use in commercial vocational schools]. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 117(3), 395–430. <https://doi.org/10.25162/zbw-2021-0018>
- Starkey, L. (2020). A review of research exploring teacher preparation for the digital age. *Cambridge Journal of Education*, 50(1), 37–56. <https://doi.org/10.1080/0305764X.2019.1625867>
- Stieff, M., Hegarty, M., & Deslongchamps, G. (2011). Identifying Representational Competence with Multi-Representational Displays. *Cognition and Instruction*, 29(1), 123–145. <https://doi.org/10.1080/07370008.2010.507318>
- Strassel, S. M., Bies, A., & Tracey, J. (2017). Situational Awareness for Low Resource Languages: The LORELEI Situation Frame Annotation Task. *SMERP@ECIR*, 32–41.
- Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2018). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods*, 50(4), 1446–1460. <https://doi.org/10.3758/s13428-017-0996-1>
- Strohmaier, A. R., Schiepe-Tiska, A., Chang, Y.-P., Müller, F., Lin, F.-L., & Reiss, K. M. (2020). Comparing Eye Movements during Mathematical Word Problem Solving in

- Chinese and German. *ZDM: The International Journal on Mathematics Education*, 52(1), 45–58. <https://doi.org/10.1007/s11858-019-01080-6>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Sweller, J., van Merriënboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2), 261–292. <https://doi.org/10.1007/s10648-019-09465-5>
- Taghipour, K., & Ng, H. T. (2016). A neural approach to automated essay scoring. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, 1882–1891.
- Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What Forty Years of Research Says About the Impact of Technology on Learning: A Second-Order Meta-Analysis and Validation Study. *Review of Educational Research*, 81(1), 4–28. <https://doi.org/10.3102/0034654310393361>
- Taub, M., & Azevedo, R. (2019). How Does Prior Knowledge Influence Eye Fixations and Sequences of Cognitive and Metacognitive SRL Processes during Learning with an Intelligent Tutoring System? *International Journal of Artificial Intelligence in Education*, 29(1), 1–28. <https://doi.org/10.1007/s40593-018-0165-4>
- Taub, M., Mudrick, N. V., Azevedo, R., Millar, G. C., Rowe, J., & Lester, J. (2017). Using multi-channel data with multi-level modeling to assess in-game performance during gameplay with Crystal Island. *Computers in Human Behavior*, 76, 641–655. <https://doi.org/10.1016/j.chb.2017.01.038>
- Telgmann, L., & Müller, K. (2023). Training & prompting pre-service teachers' noticing in a standardized classroom simulation – a mobile eye-tracking study. *Frontiers in Education*, 8, 1266800. <https://doi.org/10.3389/educ.2023.1266800>
- Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, 135(2), 77–99. <https://doi.org/10.1016/j.actpsy.2010.02.006>
- Thorne, J., Vlachos, A., Christodoulopoulos, C., & Mittal, A. (2018). *FEVER: A large-scale dataset for Fact Extraction and VERification* (arXiv:1803.05355). [arXiv. https://arxiv.org/abs/1803.05355](https://arxiv.org/abs/1803.05355)
- Thyssen, C., Huwer, J., Irion, T., & Schaal, S. (2023). From TPACK to DPACK: The “Digitality-Related Pedagogical and Content Knowledge”-Model in STEM-Education. *Education Sciences*, 13(8), 769. <https://doi.org/10.3390/educsci13080769>
- Tomczyk, L., & Fedeli, L. (2021). Digital Literacy among Teachers—Mapping Theoretical Frameworks: TPACK, DigCompEdu, UNESCO, NETS-T, DigiLit Leicester. *Proceedings of the 38th International Business Information Management Association (IBIMA)*, 244–252.
- Torres, F., Neira Tovar, L. A., & del Rio, M. S. (2017). A Learning Evaluation for an Immersive Virtual Laboratory for Technical Training Applied into a Welding Workshop. *EURASIA Journal of Mathematics, Science & Technology Education*, 13(2), 521–532.
- TPACK.org. (2023). *TPACK.ORG*. <http://tpack.org/>
- Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., & Yang, F.-Y. (2011). Visual Attention for Solving Multiple-Choice Science Problem: An Eye-Tracking Analysis. *Computers & Education*, 58(1), 375–385. <https://doi.org/10.1016/j.compedu.2011.07.012>

- Tsai, M.-J., Huang, L.-J., Hou, H.-T., Hsu, C.-Y., & Chiou, G.-L. (2016). Visual behavior, flow and achievement in game-based learning. *Computers & Education*, 98, 115–129. <https://doi.org/10.1016/j.compedu.2016.03.011>
- Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies*, 57(4), 247–262. <https://doi.org/10.1006/ijhc.2002.1017>
- Underwood, G., & Everatt, J. (1992). The Role of Eye Movements in Reading: Some Limitations of the Eye-Mind Assumption. In E. Chekaluk & K. Llewellyn (Eds.), *The Role of Eye Movements in Perceptual Processes* (pp. 111–169). Elsevier. [https://doi.org/10.1016/S0166-4115\(08\)61744-6](https://doi.org/10.1016/S0166-4115(08)61744-6)
- UNESCO. (2018). *UNESCO ICT Competency Framework for Teachers: Version 3*. UNESCO.
- Urlaub, P., & Dessen, E. (2022). Machine translation and foreign language education. *Frontiers in Artificial Intelligence*, 5, 936111. <https://doi.org/10.3389/frai.2022.936111>
- 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 Gog, T., & Jarodzka, H. (2013). Eye Tracking as a Tool to Study and Enhance Cognitive and Metacognitive Processes in Computer-Based Learning Environments. In R. Azevedo & V. Alevan (Eds.), *International Handbook of Metacognition and Learning Technologies* (pp. 143–156). Springer. https://doi.org/10.1007/978-1-4419-5546-3_10
- van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25(3), 785–791. <https://doi.org/10.1016/j.chb.2009.02.007>
- Van Gog, T., Paas, F., Marcus, N., Ayres, P., & Sweller, J. (2009). The Mirror Neuron System and Observational Learning: Implications for the Effectiveness of Dynamic Visualizations. *Educational Psychology Review*, 21(1), 21–30. <https://doi.org/10.1007/s10648-008-9094-3>
- van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2005). Uncovering expertise-related differences in troubleshooting performance: Combining eye movement and concurrent verbal protocol data: Uncovering expertise-related differences. *Applied Cognitive Psychology*, 19(2), 205–221. <https://doi.org/10.1002/acp.1112>
- van Gog, T., Paas, F., van Merriënboer, J. J. G., & Witte, P. (2005). Uncovering the Problem-Solving Process: Cued Retrospective Reporting Versus Concurrent and Retrospective Reporting. *Journal of Experimental Psychology: Applied*, 11(4), 237–244. <https://doi.org/10.1037/1076-898X.11.4.237>
- van Merriënboer, J. J. G. (2013). Perspectives on problem solving and instruction. *Computers & Education*, 64, 153–160. <https://doi.org/10.1016/j.compedu.2012.11.025>
- Van Ostaeyen, S., Embo, M., Schellens, T., & Valcke, M. (2022). Training to Support ePortfolio Users During Clinical Placements: A Scoping Review. *Medical Science Educator*, 32(4), 921–928. <https://doi.org/10.1007/s40670-022-01583-0>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 5998–6008.
- Vila, J., & Gomez, Y. (2016). Extracting business information from graphs: An eye tracking experiment. *Journal of Business Research*, 69(5), 1741–1746.

- Vrana, J., & Singh, R. (2021). Digitization, Digitalization, and Digital Transformation. In N. Meyendorf, N. Ida, R. Singh, & J. Vrana (Eds.), *Handbook of Nondestructive Evaluation 4.0* (pp. 1–17). Springer. https://doi.org/10.1007/978-3-030-48200-8_39-1
- Vrzakova, H., Begel, A., Mehtatalo, L., & Bednarik, R. (2020). Affect Recognition in Code Review: An In-situ Biometric Study of Reviewer's Affect. *Journal of Systems and Software, 159*, 110434.
- Wacker, A., Unger, V., & Rey, T. (2020). „Sind doch Corona-Ferien, oder nicht?“. Befunde einer Schüler*innenbefragung zum „Fernunterricht“ [“It's the Corona vacations, isn't it?”. Findings from a student survey on "distance learning""]. In D. Fickermann & B. Edelstein (Hrsg.), „Langsam vermiss ich die Schule ...“. *Schule während und nach der Corona-Pandemie.* (Bd. 16, S. 79–94). Waxmann. https://www.pedocs.de/front-door.php?source_opus=20230
- Wahyuni, S., Mujiyanto, J., Rukmini, D., & Fitriati, S. W. (2020). Teachers' Technology Integration Into English Instructions: SAMR Model. *Advances in Social Science, Education and Humanities Research, 443*, 546–550.
- Walstad, W. B., & Rebeck, K. (2001). Assessing the economic understanding of US High-school students. *American Economic Review, 91*(2), 452–457.
- Walstad, W. B., Rebeck, K., & Butters, R. B. (2013). The test of economic literacy: Development and results. *The Journal of Economic Education, 44*(3), 298–309.
- Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2019). GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. *arXiv:1804.07461 [Cs]*. <http://arxiv.org/abs/1804.07461>
- Wang, B., & Komatsuzaki, A. (2021, May). *GPT-J-6B: A 6 Billion Parameter Autoregressive Language Model*. <https://github.com/kingoflolz/mesh-transformer-jax>
- Wang, W., Schmidt-Crawford, D., & Jin, Y. (2018). Preservice Teachers' TPACK Development: A Review of Literature. *Journal of Digital Learning in Teacher Education, 34*(4), 234–258. <https://doi.org/10.1080/21532974.2018.1498039>
- Wang, X., Li, L., Tan, S. C., Yang, L., & Lei, J. (2023). Preparing for AI-enhanced education: Conceptualizing and empirically examining teachers' AI readiness. *Computers in Human Behavior, 146*, 107798. <https://doi.org/10.1016/j.chb.2023.107798>
- Wang, Y., Lu, S., & Harter, D. (2021). Towards Collaborative and Intelligent Learning Environments Based on Eye Tracking Data and Learning Analytics: A Survey. *IEEE Access, 9*, 137991–138002. <https://doi.org/10.1109/ACCESS.2021.3117780>
- Weinert, F. E. (1999). Die fünf Irrtümer der Schulreformer. Welche Lehrer, welchen Unterricht braucht das Land? [The five mistakes of school reformers. What teachers, what teaching does the country need?] *Psychologie Heute, 26*(7), 28–34.
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Hogrefe.
- Wheat, I. D. (2009). Teaching economics as if time mattered. In J. Reardon (Ed.), *The Handbook of Pluralist Economics Education* (pp. 69–89). Routledge.
- Wheeler, S. (2012). Distance Learning. In N. M. Seel (Hrsg.), *Encyclopedia of the Sciences of Learning* (S. 1018–1020). Springer. https://doi.org/10.1007/978-1-4419-1428-6_432
- Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Hogrefe.

- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Golemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation* [Computer software]. <https://dplyr.tidyverse.org>
- Wigand, R. T., Mertens, P., Bodendorf, F., König, W., Picot, A., & Schumann, M. (2003). *Introduction to business information systems*. Springer.
- Wilbers, K. (2009). Integrierte Unternehmenssoftware (ERP-Systeme) im kaufmännischen Unterricht [Integrated business software (ERP systems) in commercial teaching]. In H. Pongratz, P. T. Tramm, & K. Wilbers (Eds.), *Prozessorientierte Wirtschaftsdidaktik und Einsatz von ERP-Systemen im kaufmännischen Unterricht* (Vol. 4, pp. 61–76). Shaker.
- Wilbers, K. (2014). *Wirtschaftsunterricht gestalten [Designing business education]*. epubli. https://www.pedocs.de/frontdoor.php?source_opus=8420
- Williams, A., Nangia, N., & Bowman, S. R. (2018). *A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference* (arXiv:1704.05426). arXiv. <http://arxiv.org/abs/1704.05426>
- Williamson, D. M., Mislevy, R. J., & Bejar, I. I. (2006). Automated scoring of complex tasks in computer-based testing: An Introduction. In D. M. Williamson, R. J. Mislevy, & I. I. Bejar (Eds.), *Automated scoring of complex tasks in computer-based testing* (pp. 1–13). Psychology Press.
- Wisiecka, K., Krejtz, K., Krejtz, I., Sromek, D., Cellary, A., Lewandowska, B., & Duchowski, A. (2022). Comparison of Webcam and Remote Eye Tracking. *2022 Symposium on Eye Tracking Research and Applications*, 1–7. <https://doi.org/10.1145/3517031.3529615>
- Wößmann, L., Freundl, V., Grewenig, E., Lergetporer, P., Werner, K., & Zierow, L. (2020). Bildung in der Coronakrise: Wie haben die Schulkinder die Zeit der Schulschließungen verbracht, und welche Bildungsmaßnahmen befürworten die Deutschen? [Education during the coronavirus crisis: How have schoolchildren spent the period of school closures and what educational measures do Germans support?] *ifo Schnelldienst*, 73(09), 25–39.
- Wößmann, L., Freundl, V., Grewenig, E., Lergetporer, P., Werner, K., & Zierow, L. (2021). Bildung erneut im Lockdown: Wie verbrachten Schulkinder die Schulschließungen Anfang 2021? [Education in lockdown again: how did schoolchildren spend the school closures at the start of 2021?] *ifo Schnelldienst*, 74(5), 36–52.
- Wu, C.-H., Tzeng, Y.-L., & Huang, Y. M. (2014). Understanding the relationship between physiological signals and digital game-based learning outcome. *Journal of Computers in Education*, 1(1), 81–97. <https://doi.org/10.1007/s40692-014-0006-x>
- Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, Ł., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., ... Dean, J. (2016). Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation. *arXiv:1609.08144 [Cs]*. <http://arxiv.org/abs/1609.08144>
- Wuttke, E., Seifried, J., & Niegemann, H. M. (Eds.). (2020). *Vocational education and training in the age of digitization: Challenges and opportunities*. Barbara Budrich.
- Wuttke, E., Seifried, J., & Niegemann, H. (2020). Editorial: Vocational Education and Training in the Age of Digitization—Challenges and Opportunities. In E. Wuttke, J. Seifried, &

- H. Niegemann (Eds.), *Vocational Education and Training in the Age of Digitization: Challenges and Opportunities* (pp. 15–19). Barbara Budrich. <https://doi.org/10.2307/j.ctv18dvv1c>
- Wyss, C., Rosenberger, K., & Bühner, W. (2020). Student Teachers' and Teacher Educators' Professional Vision: Findings from an Eye Tracking Study. *Educational Psychology Review*, 33, 91–107. <https://doi.org/10.1007/s10648-020-09535-z>
- Xie, H., Zhao, T., Deng, S., Peng, J., Wang, F., & Zhou, Z. (2021). Using eye movement modelling examples to guide visual attention and foster cognitive performance: A meta-analysis. *Journal of Computer Assisted Learning*, 37(4), 1194–1206. <https://doi.org/10.1111/jcal.12568>
- Yamamoto, M., Umemura, N., & Kawano, H. (2018). Automated Essay Scoring System Based on Rubric. In R. Lee (Ed.), *Applied computing & information technology* (pp. 177–190). Springer. https://doi.org/10.1007/978-3-319-64051-8_11
- Yang, F.-Y., Tsai, M.-J., Chiou, G.-L., Lee, S. W.-Y., Chang, C.-C., & Chen, L.-L. (2018). Instructional suggestions supporting science learning in digital environments based on a review of eye tracking studies. *Journal of Educational Technology & Society*, 21(2), 28–45.
- Yang, X., Zhang, L., & Yu, S. (2017). Can Short Answers to Open Response Questions Be Auto-Graded Without a Grading Rubric? In E. André, R. Baker, X. Hu, Ma. M. T. Rodrigo, & B. du Boulay (Eds.), *Artificial Intelligence in Education (AIED)* (pp. 594–597). Springer. https://doi.org/10.1007/978-3-319-61425-0_72
- Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R., & Le, Q. V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in Neural Information Processing Systems*, 32.
- Yin, W., Hay, J., & Roth, D. (2019). Benchmarking Zero-shot Text Classification: Datasets, Evaluation and Entailment Approach. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 3912–3921. <https://doi.org/10.18653/v1/D19-1404>
- Zancajo, A., Verger, A., & Bolea, P. (2022). Digitalization and beyond: The effects of Covid-19 on post-pandemic educational policy and delivery in Europe. *Policy and Society*, 41(1), 111–128. <https://doi.org/10.1093/polsoc/puab016>
- Zetland, D., Russo, C., & Yavapolkul, N. (2010). Teaching Economic Principles: Algebra, Graph or Both? *The American Economist*, 55(1), 123–131. <https://doi.org/10.1177/056943451005500113>
- Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence*, 2, 100025. <https://doi.org/10.1016/j.caeai.2021.100025>
- Zhang, W., & Tang, J. (2021). Teachers' TPACK Development: A Review of Literature. *Open Journal of Social Sciences*, 9(7), 367–380. <https://doi.org/10.4236/jss.2021.97027>
- Zhang, X., Zhao, J., & LeCun, Y. (2015). *Character-level Convolutional Networks for Text Classification* (Version 3). arXiv. <https://doi.org/10.48550/ARXIV.1509.01626>
- Zhong, R., Lee, K., Zhang, Z., & Klein, D. (2021). Adapting language models for zero-shot learning by meta-tuning on dataset and prompt collections. *arXiv Preprint arXiv:2104.04670*.

- Zhuang, Q., & Liu, X. (2022). Comparison of graph and animation: An unbalanced battle over two decades. *Frontiers in Psychology*, *13*, 810557. <https://doi.org/10.3389/fpsyg.2022.810557>
- Zinke, G. (2019). *Berufsbildung 4.0 - Fachkräftequalifikationen und Kompetenzen für die digitalisierte Arbeit von morgen: Branchen- und Berufscreening: vergleichende Gesamtstudie [Vocational education and training 4.0 - Skilled qualifications and competences for the digitalized work of tomorrow: Sector and occupation screening: comparative overall study]*. Bundesinstitut für Berufsbildung.
- Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver-Operating Characteristic Analysis for Evaluating Diagnostic Tests and Predictive Models. *Circulation*, *115*(5), 654–657. <https://doi.org/10.1161/CIRCULATIONAHA.105.594929>
- Zutavern, S., & Seifried, J. (2022). Vocational Education and Training in Germany: Benefits and Drawbacks of the Dual Approach as Preparation for Professional Employment. In C. Harteis, D. Gijbels, & E. Kyndt (Eds.), *Research Approaches on Workplace Learning* (pp. 347–365). Springer. https://doi.org/10.1007/978-3-030-89582-2_16

Appendix

Appendix A: Study 3 (Chapter 5). Analysing domain-specific problem-solving processes within authentic computer-based learning and training environments by using eye-tracking: A scoping review.

Table 5-4: Descriptive overview of included studies.

Reference	Domain	Sample	Task(s)	Performance measures
Abele et al. 2017	Engineering	15 apprentices in automotive mechatronics engineering from vocational education schools in Germany	Troubleshooting two scenarios in a computer-based motor vehicle simulation.	Manual scoring (number of valid errors scored)
Cloude et al. 2020	Microbiology	68 students from three universities in the USA	Gathering evidence, testing hypotheses (scientific thinking) to solve a mysterious disease outbreak in Crystal Island, an open-world 3D serious game.	Pretest-Posttest
Dubovi 2022	Healthcare	61 nursing students from a university in Israel	Collecting clinical data in a virtual hospital and conducting basic medication procedures to solve the complex problem of identifying the correct medication.	Pretest-Posttest
Emerson et al. 2020	Microbiology	61 students from three universities in the USA	Gathering evidence, testing hypotheses (scientific thinking) to solve a mysterious disease outbreak in Crystal Island, an open-world 3D serious game.	Pretest-Posttest
Gomes et al. 2013	Engineering	7 high school students from the USA	Solving engineering problems in three short games.	Automated Scoring (solve as many levels as possible in a limited time)
Kang & Landry, 2014	Air Traffic Control	60 undergraduate and graduate engineering students ($n = 20$ for treatment group) from the USA	Solving possible air traffic conflicts by detecting a loss of separation among multiple aircraft and multiple types of conflicts after training by following an expert scan path.	Manual Scoring (number of correct detection/false alarms)

Reference	Domain	Sample	Task(s)	Performance measures
Lee et al. 2019	Medicine	24 professionals and 22 medical students ($n = 46$) from the Netherlands	Stabilising a virtual patient by applying the ABCDE method within a maximum of 15 minutes in an authentic computer-based medical simulation game.	Automated Scoring (Intervention completion score by extracting data from log files)
Lee et al. 2020	Medicine	70 medical students from the Netherlands	Stabilising a virtual patient within a maximum of 15 minutes in an authentic computer-based medical simulation game.	Automated Scoring (log files)
Sohn et al. 2005	Military	12 students and staff from the USA	Learning subtasks (identification and control of information search) of tactical decision-making in an anti-air-warfare simulation.	Manual Scoring (Immediate feedback on timing and accuracy)
Taub et al. 2017	Microbiology	50 non-biology students from the USA.	Gathering evidence, testing hypotheses (scientific thinking) to solve a mysterious disease outbreak in Crystal Island, an open-world 3D serious game.	Automated Scoring (log files, in-game assessments as self-regulated judgments of learning)
Tsai et al. 2016	Electrophysics	22 university students from Taiwan	Solving a problem-based learning task embedded in a game-based scenario by gathering information and applying knowledge about electromagnets to escape from a lab at the end.	Pretest-Posttest
(van Gog et al. 2005a)	Electronics	10 students from the Netherlands	Troubleshooting malfunctions in electrical circuits in a simulation.	Manual Scoring (on-task solution aspects)

Table 5-5: Overview of research questions/hypotheses and main findings.

Reference	Research Questions / Hypotheses	Main findings
Abele et al. 2017	Successful subjects show a higher total critical fixation duration than less successful subjects.	<ul style="list-style-type: none"> - higher performance → longer total fixation time - lower performance → shorter total fixation time or substantial longer fixation time - more prolonged fixation might indicate confusion at some point
Cloude et al. 2020	<p>Time fixating and interacting with scientific reasoning-related game elements predicts post-test scores (RQ1).</p> <p>Time fixating on scientific reasoning-related game elements predicts time interacting with scientific reasoning-related game elements (RQ2).</p> <p>Time fixating on non-scientific reasoning-related game elements predicts the time interacting with non-scientific reasoning-related game elements (RQ3).</p>	<ul style="list-style-type: none"> - lower performance → higher proportion of time gathering information and less time generating hypotheses - lower prior knowledge positively moderates the relation between interaction and fixation on gathering information in the GBLE while a negative relation for higher prior knowledge was found - no relation between interaction and fixation for non-scientific reasoning was found - eye-tracking data helps to supplement and contextualize log-files - multichannel data may be needed for individualized learning analytics approaches
Dubovi 2022	Students' cognitive and emotional engagement can be measured by their self-reports and psycho-physiological real-time measurements and the synergistic effect of cognitive and emotional engagement on learning.	<ul style="list-style-type: none"> - emotional and cognitive engagement via multimodal metrics explained 51% of post-test learning achievement - no significant impact of joy expression on post-test - frequent anger expressions were associated with lower post-test scores - no significant change in self-reported affective state over three times - blink rate is negatively associated with post-test scores and shows significantly lower rates during the actual problem - more visual attention is spent on similar medicine, indicating processing difficulties through fixation counts and dwells - significant correlation between EDA peaks and blinks but not with emotional engagement - positive emotions were related to inducing blinks - a higher level of presence was related to more visual attention to relevant medicine

Reference	Research Questions / Hypotheses	Main findings
Emerson et al. 2020	<p>Student gameplay behavior traces, facial expressions of emotions and eye gaze classify low, medium, and high performing groups. (RQ1)</p> <p>Student gameplay behavior traces, facial expressions of emotions and eye gaze classify low, medium, and high-interest groups. (RQ2)</p>	<ul style="list-style-type: none"> - gaze as a feature (unimodal) or gameplay + face as a multimodal feature approach yields in the accuracy of .67 for prediction among three performance groups - gameplay + face (multimodal) yields in .59 accuracy for prediction among three interest level groups - adding more modalities comes at the cost of noise, so feature selection must be done carefully to avoid overfitting
Gomes et al. 2013	Differences in eye-tracking patterns exist between students with low and high performance in the three engineering-related computer games.	<ul style="list-style-type: none"> - shorter time to first fixation, fewer clicks, more unique fixations, and a longer duration per fixation for high-performance cluster - longer time for the first fixation, a higher number of clicks and short fixation durations might indicate a lack of focus on strategy or reasoning before action (“trial-and-error”) - shorter durations for first fixations might indicate higher attentional readiness and indicates more time spent on reasoning before action
Kang & Landry, 2014	The performance will be different for novices exposed to the expert scan path compared to the control group or novices without treatment.	<ul style="list-style-type: none"> - treatment (expert scan path) group showed significantly fewer false alarms than the verbal instruction group or control group - treated novices tend to follow a professional expert scan pattern after treatment (circular) - treatment group perceived expert scan paths as helpful, and a scan path could improve the training of novices

Reference	Research Questions / Hypotheses	Main findings
Lee et al. 2019	<p>Participants with high domain-specific prior knowledge (DSPK, i.e., experts) show higher systematicity in approach than participants with low DSPK (i.e., novices). (H1)</p> <p>Experts show higher accuracy in visual selection by allocating more visual attention to critical diagnosis areas (H2a) and in motor reactions by completing more interventions (H2b) than novices.</p> <p>Experts show higher speed in performance by completing interventions faster than novices. (H3)</p> <p>Experts experience lower cognitive load than novices. (H4)</p>	<p>- experts-Novice comparison shows for experts: → more systematicity (indicated by HMM score) → higher proportions of dwell time to total time (large effect), a higher ratio of fixation count to total fixation counts (medium effect), and longer fixation duration (large effect) on critically relevant information → no difference for other AOIs except for the intervention area; a lower proportion of total fixation counts (medium effect) → no difference for average fixation duration and fixation count, but cognitive load and transition rate correlate negatively with self-reported NASA-TLX score.</p>
Lee et al. 2020	<p>Cognitive load in the pause-available condition (PA) would be higher than in the pause-unavailable condition (PU), at the overall level. (H1a)</p> <p>Performance in the PA would be higher than in the PU, at the overall level. (H1b)</p> <p>Within PA, cognitive load in the pause-taking group (PAn) would be lower than in the no-pause-taking group (PA0), at the overall level. (H2a)</p> <p>Within the PA, performance in the PAn would be higher than in the PA0, at the overall level. (H2b)</p> <p>In the absence of intense events, the cognitive load would increase during pauses. (H3a)</p> <p>In an intense situation, the cognitive load would decrease during pauses. (H3b)</p>	<p>- overall, allowing pauses increases performance and cognitive load, regardless of whether pauses were taken or not. - when pauses were available, taking those pauses did not further benefit cognitive load or performance. - during pauses cognitive load was lower compared to simulation. - pupillometry might be a valid measure of the cognitive load next to self-reports.</p>
Sohn et al. 2005	<p>Participants learn to pay more attention to task-relevant regions and less attention to task-irrelevant regions with practice over time.</p>	<p>- information-seeking behaviour changed over time; reduction in time on relevant and irrelevant regions.</p>

Reference	Research Questions / Hypotheses	Main findings
Taub et al. 2017	<p>The more books' participants read, and the more often they read each book, the fewer concept matrix submission attempts they made, resulting in better performance. (H1)</p> <p>The longer fixation durations on the book content and concept matrices, the fewer concept matrix attempts, resulting in better performance. (H2)</p> <p>There will be a significant interaction, such that log file data (number of books and frequency of reading each book) and eye-tracking data (proportions of fixations on book content and book concept matrices) will jointly impact concept matrix submission attempts, with higher levels of all variables resulting in fewer attempts, and thus greater performance. (H3)</p>	<p>- negative effect between number of books and performance as well as for frequency of books and performance</p> <p>- but the best performance was associated with fewer books and higher frequencies per book</p> <p>- reading more books (quantity) might not improve performance while reading books several times (quality) might do</p> <p>- no unique association between proportions of fixations on book content or book concept matrix with submission attempts were found, but a significant interaction effect. Low proportions of fixations on book content and concept matrices were related to high performance</p> <p>- significant associations between performance and the multimodal predictors as well as for the interaction term. The highest performance was related to a higher frequency of books, fewer books, and lower proportions of fixations on book content or concept matrix</p>
Tsai et al. 2016	<p>Do players with different conceptual comprehension in GBL: - have different visual attention distributions while playing games? If yes, what are the patterns for high and low-achievement players? (RQ1)</p> <p>- have different patterns of visual attention transactions (representing the players' control strategies of multi-tasking coordination applied in the game)? (RQ2)</p> <p>- experience different levels of game flow? (RQ3)</p>	<p>low comprehension group:</p> <p>- higher PFD and PCD in the components area</p> <p>- higher mental effort (heatmap)</p> <p>- viewed graphical information more frequently than the high comprehension group</p> <p>- paid more attention to graphic information according to heatmap analyses (while the high comprehension group spent less attention on graphical and more attention on textual information)</p> <p>- low comprehension group tended to get stuck in the message (cues) and out-of-screen gaze while the successful group tend to transfer knowledge and might use out-of-screen gaze as a pausing/reasoning strategy.</p> <p>high comprehension group showed</p> <p>- a higher sense of control and concentration, flow experience and visual attention association: higher flow time distortion → fixations on the main task</p> <p>lower flow time distortion → fixations on the message prompt</p>

Reference	Research Questions / Hypotheses	Main findings
(van Gog et al. 2005a)	<p>Higher expertise participants spend more time on problem orientation, problem formulation, deciding on actions and evaluating them, while lower expertise participants are more likely to test out the functioning of the circuit to try to generate new hypotheses. (1)</p> <p>Higher expertise participants' orientation and evaluation phase will be less cognitively demanding than reasoning, and all these processes are more demanding for lower than higher expertise participants. In the 'problem orientation' phase, higher expertise participants will have a higher proportion of fixations on components related to major faults. (2)</p> <p>Eye movement and concurrent verbal protocol data together show how eye movement data may make to the investigation of cognitive processes. (3)</p>	<p>- high expertise participants spent more time in the 'problem orientation' and 'action evaluate & next action decision' phase (but not for the 'problem formulation' phase)</p> <p>- higher expertise groups only differ for mean fixation duration over all phases, but show more fixations on fault-related components, they show:</p> <p>→ shorter mean fixation duration in the 'orientation' phase</p> <p>→ longer mean fixation duration in the 'problem formulation' phase</p> <p>→ verbal data that reveals predictive behaviour</p> <p>- Low expertise participants' verbal data show no orientation and an unstructured initial testing approach</p>

Appendix B: Study 4 (Chapter 6). Visual attention while solving economic learning tasks using dynamic, interactive graphs – An eye-tracking study.

Table 6-6: Descriptive statistics for tasks with static graphs.

Variable	<i>M</i>	<i>SD</i>	1	2	3
1. Performance (static graphs)	3.90	1.62			
2. Prior knowledge	5.16	1.87	.40**		
3. Attitude	2.98	0.49	.24	.16	
4. Total Fixation Duration ^a	0.00	1.00	-.09	-.32*	-.16

Note. ^a standardized. * indicates $p < .05$. ** indicates $p < .01$.

Table 6-7: Descriptive statistics for tasks with dynamic graphs.

Variable	<i>M</i>	<i>SD</i>	1	2	3
1. Performance (dynamic graphs)	4.18	1.55			
2. Prior knowledge	5.16	1.87	.43**		
3. Attitude	2.98	0.49	.32*	.16	
4. Total Fixation Duration ^a	-0.00	1.00	.22	.04	-.23

Note. ^a standardized. * indicates $p < .05$. ** indicates $p < .01$.

Task 1 (easy exercise) was instructed as follows: ‘You have already learned a lot about how markets work during your studies. Now it's time to apply what you've learned. Food is also traded in markets. Therefore, the supply and demand model is a good one to apply here. Exercise: Let's take the market for cucumbers. There has been a crop failure due to very cold weather. You are now asked to predict how this will affect supply and demand, as well as the quantity of cucumbers sold on the market, the market price, producers' surplus, and consumers. An example single-choice question involves the supply curve, which can shift left, shift right, or stay the same. Another example item relates to market prices, which can increase, stay the same, or decrease. Task 2 (difficult exercise) was instructed as follow: ‘It is not always the case

that the equilibrium price and quantity arise freely in the market. In some cases, the government interferes with market mechanisms. For example, the government may set maximum or minimum prices for a market. An example of this is the minimum wage in Germany. In the following, we will look at the effects of a maximum price on the market situation as an example. Exercise: You may be familiar with this problem. You want to rent an apartment, but the prices are extremely high. There seems to be a simple solution to this problem, doesn't there? The government should set maximum prices so that landlords cannot charge exorbitant rents. The government follows this idea and wants to help the tenants. To this end, it sets a maximum price for apartments of a certain size, which is below the market price. How will this affect the market situation?' An example single-choice question refers to effect on the sum of producers' and consumers' surplus, which can increase, remain unchanged, or decrease. Students had to select the correct answer using a static graph and, in another scenario, a dynamic graph.

Appendix C: Study 5 (Chapter 7). Prompt text classifications with transformer models! An exemplary introduction to prompt-based learning with large language models.

- Supplementary Material: <https://github.com/LucaOffice/Publications>
- Funding: This research was funded by the German Federal Ministry of Education and Research, grant number 01DB1119-23.
- Institutional Review Board Statement: The study was approved by the Institutional Review Board (or Ethics Committee) of University of Bamberg, Germany.
- Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.
- Data Availability Statement: The dataset including the full texts of the response emails is unfortunately not publicly available.
- Conflicts of Interest: The authors declare no conflict of interest.
- BART-large-mnli labels provided for zero-shot pipeline:
candidate_labels = ['unhöflich', 'höflich']
- GPT-J-GB few-shot prompt:

few_shot = ("""Klassifizier die Mail. Labels: höflich, unhöflich

#####

mail: Hallo Susanne, unter gegebenen Umständen kann ich die folgenden Aufgaben nicht bearbeiten. Die Excel-Simulation ist nicht dafür geeignet. Darüber hinaus funktioniert mein Taschenrechner nicht. Mit freundlichen Grüßen
Formulierung ist höflich

#####

mail: bitte kommen Sie rüber und erklären mir es selbst
Formulierung ist unhöflich

#####

mail: Guden, Ich habe keinerlei Ahnung ;) Gruß der Boss!
Formulierung ist unhöflich

#####

mail: Flexi
Formulierung ist unhöflich

#####

mail: Sehr geehrte Frau Meier, ich habe die von Ihnen gestellte Aufgabe leider nicht lösen können, da ich im Umgang mit solchen Aufgaben noch nicht so viel Erfahrung habe.

Formulierung ist höflich

#####

""")

Declaration in Lieu of Oath

Declaration in lieu of oath according to section 8 subsection 2 No. 1(b) of the Regulations and Procedures Governing the Doctoral Dissertation to Earn a Doctoral Degree in Business at the University of Mannheim

Eidesstattliche Versicherung

Eidesstattliche Versicherung gemäß § 8 Absatz 2 Satz 1 Buchstabe b) der Promotionsordnung der Universität Mannheim zur Erlangung des Doktorgrades der Betriebswirtschaftslehre (Dr. rer. pol.)

1. The submitted doctoral dissertation on the subject ‘*Digitalization in Economic and Business Education – Exploring how technology can support and enhance educational processes*’ is my own work and to the rules of proper scientific conduct.

Bei der eingereichten Dissertation mit dem Titel ‘Digitalization in Economic and Business Education – Exploring how technology can support and enhance educational processes’ handelt es sich um mein eigenständig erstelltes Werk, das den Regeln guter wissenschaftlicher Praxis entspricht.

2. I did not seek unauthorized assistance of a third party and I have employed no other sources or means except the ones listed. I clearly marked any direct and indirect quotations derived from the works of others.

Ich habe nur die angegebenen Quellen und Hilfsmittel benutzt und mich keiner unzulässigen Hilfe Dritter bedient. Insbesondere habe ich wörtliche und nicht wörtliche Zitate aus anderen Werken als solche kenntlich gemacht.

3. I did not yet present this doctoral dissertation or parts of it at any other higher education institution in Germany or abroad.

Die Arbeit oder Teile davon habe ich bislang nicht an einer Hochschule des In- oder Auslands als Bestandteil einer Prüfungs- oder Qualifikationsleistung vorgelegt.

4. I hereby confirm the accuracy of the affirmation above.

Die Richtigkeit der vorstehenden Erklärung bestätige ich.

5. I am aware of the significance of this affirmation and the legal ramifications in case of untrue or incomplete statements. I affirm in lieu of oath that the statements above are to the best of my knowledge true and complete.

Die Bedeutung der eidesstattlichen Versicherung und die strafrechtlichen Folgen einer unrichtigen oder unvollständigen eidesstattlichen Versicherung sind mir bekannt. Ich versichere an Eides statt, dass ich nach bestem Wissen die reine Wahrheit erklärt und nichts verschwiegen habe.

I agree that for the purpose of assessing plagiarism the dissertation may be electronically forwarded, stored, and processed.

Ich bin damit einverstanden, dass die Arbeit zum Zwecke des Plagiatsabgleichs in elektronischer Form versendet, gespeichert und verarbeitet wird.

Mannheim,

Christian Mayer

Place, Date

Signature

Ort, Datum

Unterschrift

Author's addendum: During the preparation of this work, the author used ChatGPT, DeepL, Grammarly and similar large language models in order to optimize the language style. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the content of the publication.

Ergänzung des Autors: Bei der Vorbereitung dieser Arbeit hat der Autor ChatGPT, DeepL, Grammarly und ähnliche große Sprachmodelle verwendet, um den Sprachstil zu optimieren. Nach der Nutzung dieser Tools/Dienste hat der Autor den Inhalt nach Bedarf überprüft und bearbeitet und übernimmt die volle Verantwortung für den Inhalt der Veröffentlichung.

Doctoral Study Program

This thesis resulted from individual doctoral studies at the

University of Mannheim

Business School

Area Economic and Business Education

Chair of Professional Teaching and Learning

which were complemented by the successful completion of a doctoral program at the

Graduate School of Economic and Social Sciences (GESS)

OpenCampus.Sh

which included the following courses:

1. OpenCampus.Sh Machine Learning Degree: 17.5 ECTS
Dr. Steffen Brandt
 1. *Spring 20 - Spring 22, several courses and hackathon participation, Grade: passed.*
2. GESS MAN 805: Applied Methods in Management Research, 6 ECTS
Prof. Dr. Torsten Biemann
 2. *Fall 22, Grade: 1.7*
3. GESS IS 809: Advanced Data Science Lab II (Text Mining), 6 ECTS
Prof. Dr. Markus Strohmaier
 3. *Spring 23, Grade: 1.3*

Curriculum Vitae

Christian W. Mayer

Employment

since 2020 Research assistant, Chair of Economic and Business Education, Professional Teaching and Learning (Prof. Jürgen Seifried), University of Mannheim, Germany

Education

since 2020 Doctoral Study Program, University of Mannheim
 2017 – 2020 Master of Science Business Education, University of Mannheim
 2014 – 2017 Bachelor of Science Business Education, University of Mannheim

Selected Teaching activities

since 2021 Design of economic and vocational learning environments II – Graduate level, German, spring and fall semester
 Current didactic issues: Teaching and Learning with Enterprise Resource Planning Systems - SAP4school, Graduate level, German, spring and fall semester
 Supervising bachelor theses and master theses

Selected Publications

- 2023 **Mayer, C. W.**, Ludwig, S. & Brandt, S. (2023). Prompt text classifications with transformer models! An exemplary introduction to prompt-based learning with large language models. *Journal of Research on Technology in Education: JRTE*, 55(1), 125–141. <https://doi.org/10.1080/15391523.2022.2142872>
- Mayer, C. W.**, Rausch, A. & Seifried, J. (2023). Analysing domain-specific problem-solving processes within authentic computer-based learning and training environments by using eye-tracking: a scoping review. *Empirical Research in Vocational Education and Training*, 15:2, 1–27. <https://doi.org/10.1186/s40461-023-00140-2>
- 2022 **Mayer, C. W.**, (2022). Lehren und Lernen mit Enterprise Resource Planning (ERP) Systemen – Typische Bearbeitungsprobleme als Grundlage der Lernprozessgestaltung. *Berufs- und Wirtschaftspädagogik Online:bwp@*, 43, 1–17. https://www.bwpat.de/ausgabe43/mayer_bwpat43.pdf
- Gorshid, G. D., **Mayer, C.**, Rausch, A. & Seifried, J. (2022). Das LUCA-Dashboard im Usability-Test – Eine gaze-cued retrospective Think-Aloud-Studie. In S. Schumann, S. Seeber, S. A-

bele (Eds.), *Digitale Transformation in der Berufsbildung: Konzepte, Befunde und Herausforderungen* (Band 41, 189–212). Bielefeld: wbv.

- 2021 Ludwig, S., **Mayer, C.**, Hansen, C., Eilers, K. & Brandt, S. (2021). Automated essay scoring using transformer models. *Psych*, 3(4), 897–915. <https://doi.org/10.3390/psych3040056>
- Seifried, J., Gentner, S., Brandt, S., Braunstein, A., Deutscher, V., Gorshid, G. D., Ludwig, S., **Mayer, C.**, Rausch, A. and Winther, E. (2021). Flexibel einsetzbare Lehr- und Lernplattform für kaufmännische Bildung: LUCA Office Simulation. *VLB-Akzente: berufliche Bildung in Bayern*, 30, 9–11.

Selected Presentations

- 2024 **Mayer, C.**, Mühldorfer, R. & Seifried, J. (2024, April). *Perceptions of immersion in a student-teacher course using an immersive virtual reality (VR) classroom*. Roundtable presentation at the 2024 Annual Meeting of the AERA. Philadelphia, USA.
- 2023 **Mayer, C.**, Findeisen, S., Guggemos, J. & Seifried, J. (2023, September). *Visuelle Aufmerksamkeit beim Lösen von ökonomischen Single-Choice-Aufgaben mit dynamischen Grafiken - Eine Eye-Tracking-Studie*. Paper presented at the Jahrestagung der Sektion Berufs- und Wirtschaftspädagogik 2023 der DGFE, Flensburg, Germany.
- 2022 **Mayer, C.** & Seifried, J. (2022, August). *Learning and teaching with Enterprise Resource Planning (ERP) systems – A report on self-reflected typical student mistakes and problem-solving approaches by student teachers*. Paper presented at the 11th EARLI SIG14 Conference 2022, Paderborn, Germany.
- 2021 **Mayer, C.**, Gentner, S., Seifried, J., Deutscher, V. & Rausch, A. (2021, September). *Digitaler Unterricht an kaufmännischen Schulen – Eine Charakterisierung der aktuellen Unterrichtspraxis auf Basis des SAMR-Modells*. Paper presented at the Jahrestagung der Sektion Berufs- und Wirtschaftspädagogik der DGfE, Bamberg (online), Germany.
- 2020 **Mayer, C.** & Seifried, J. (2020, September). *Eye-tracking zur Analyse des kaufmännischen Problemlöseprozesses in einer computergetriebenen Office Simulation*. Paper presented at the Jahrestagung der Sektion Berufs- und Wirtschaftspädagogik der DGfE, Osnabrück (online), Germany.