

Master Thesis

Understanding the potential of Augmented Reality in manufacturing processes

Felix Kaufmann – 5123585

August 2021



Understanding the potential of Augmented Reality in manufacturing processes

Master thesis submitted to Delft University of Technology
in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

In Management of Technology

Faculty of Technology, Policy and Management

by

Felix Kaufmann

Student number: 5123585

To be defended in public on August 27th 2021

Graduation committee

Chairperson : Prof. dr., Frances Brazier, Multi Actor Systems
First Supervisor : Dr., Laurens Rook, Value, Technology and Innovation
Second Supervisor : Dr., Iulia Lefter, Multi Actor Systems
External Supervisor : Martin Pätzold, Henkel AG & Co. KGaA, Engineering & Maintenance

Abstract

Increasing flexibility requirements and skill gaps resulting from today's world of globalisation and digitisation pose constant challenges for manufacturing companies. Augmented Reality (AR) applications offer an efficient way to overcome these tensions by enhancing the interaction between people and technology. However, individual models in the scientific literature show ambiguous findings, and a statistically powerful empirical assessment is still missing.

Hence, this research project aims to understand the potential of AR applications in manufacturing environments by aggregating the empirical findings. For this purpose, the following research question is posed: 'Can the use of AR solutions benefit manufacturing activities and if so, how?'. Following the media naturalness theory by Kock [2005], this research hypothesises that AR solutions in comparison to classical instructions have a reducing effect on processing times, errors rates, and cognitive load levels of workers during manufacturing activities.

To answer the research question and prove the hypotheses, this research project conducts three meta-analyses in which several small studies are synthesised into one large study. Specifically, the meta-analyses address the evaluation criteria 'time', 'errors', and 'cognitive load'. The underlying systematic literature search to collect and evaluate relevant data follows the framework by Vom Brocke et al. [2009]. What is more, this research project examines the inter-relationships between the evaluation criteria and moderating variables using meta-regressions. Finally, surveys with industrial experts in a consumer goods and chemical company support and expand the findings from the meta-analyses and the meta-regressions.

The meta-analyses show that AR applications in comparison to classical instructions indeed have a reducing effect on the described evaluation criteria. In particular, based on the studies, a small, reducing effect can be achieved for 'time', a medium, reducing effect for 'errors', and a small to medium, reducing effect for 'cognitive load'. For this reason, all three previously formulated hypotheses are accepted. Furthermore, in line with the media naturalness hypothesis by Kock [2005, p. 122], the meta-regressions show that 'cognitive load' moderates the evaluation criterion 'time'. The results are validated with the help of the expert surveys in the company context, with time savings being identified as the greatest potential and lack of proven profitable business models as the greatest challenge.

Further research could, on the one hand, focus on repeating the meta-analyses as soon as new empirical studies are available and on the analysis of moderating variables. On the other hand, a long-term validation in manufacturing environments across industries is still missing and could show further scientific and practical relevance.

Acknowledgements

Take risks: If you win, you will be happy. If you lose, you will be wise. - Unknown

Despite all the challenges as a result of the current pandemic, I am more than happy that I took the risk and started the master program Management of Technology at Delft University of Technology. Besides knowledge and analytical skills, the most important gain from my time in Delft are the (hopefully everlasting) friendships with smart and inspiring fellow students and housemates from all around the world. The last two years have been a great opportunity to step out of my comfort zone and to develop my personal interests, ambitions and preferences.

A large part of my journey in Delft presents the completion of this master's thesis. I have always been enthusiastic about new technologies, especially in the field of manufacturing. Hence, it was a great experience to support the digitisation of the factories at Henkel and better understand the influence of AR in such environments. I am more than happy that I have been part of the Henkel family for six months and appreciate the continuous support by my colleagues.

I would like to thank my first supervisor Dr. Laurens Rook for all his support during the last months. I highly appreciate every helpful suggestion and constructive feedback during the preparation of this thesis. Without your support I would have never been given the chance to present my research ideas at the NeuroIS 2021. I am especially proud of our success to publish and present the initial ideas of this thesis.

Along the way, Prof. Frances Brazier and Dr. Iulia Lefter were a great support as well and I would like to express my special thanks to you. I am very grateful for your precious input and inspiring criticism which improved the outcome of the conference paper, conference presentation, and this thesis with no doubt.

Ultimately, my deepest appreciation goes to my parents Madeleine and Jens, sister Tiffany, grandmother Ingeborg, and beloved friends. Thanks to you, I have been relatively successful in this long journey, from my first day in the bachelor's program to my soon-to-be master's degree. An adventurous and exciting time is ahead and I am already looking forward to continue our paths together.

Contents

- 1 Introduction** **1**
 - 1.1 Motivation 1
 - 1.2 Research Objective and Questions 2
 - 1.3 Report Structure 3

- 2 Literature Review** **5**
 - 2.1 Manufacturing Environments 5
 - 2.1.1 Relevance of Information in Manufacturing Environments 5
 - 2.2 Augmented Reality in Manufacturing Environments 7
 - 2.2.1 Definition Augmented Reality 7
 - 2.2.2 Fields of Application Augmented Reality in Manufacturing Environments 8
 - 2.3 Existing Approaches 9
 - 2.3.1 Relevant Factors and Measurement 10
 - 2.4 Knowledge Gap 11

- 3 Hypothesis Development** **15**
 - 3.1 Definition Hypothesis 15
 - 3.2 Derivation of Hypothesis 15

- 4 Methodological Approach** **19**

- 5 Meta-Analysis** **23**
 - 5.1 Definition Meta-Analysis 23
 - 5.1.1 Fields of Application Meta-Analysis 23
 - 5.1.2 Requirements Meta-Analysis 24
 - 5.1.3 Methodological Approach Meta-Analysis 25
 - 5.1.4 Advantages Meta-Analysis 30
 - 5.1.5 Weaknesses and Sources of Error Meta-Analysis 30
 - 5.2 Execution Meta-Analysis 33
 - 5.2.1 Formulation of a Research Question 33
 - 5.2.2 Collection of Data 34
 - 5.2.3 Evaluation of Data 36
 - 5.2.4 Analysis and Interpretation of the Data 41
 - 5.2.5 Performance Sensitivity Analysis 46
 - 5.3 Summary Meta-Analysis 51

- 6 Meta-Regression** **53**
 - 6.1 Definition Meta-Regression 53
 - 6.2 Execution Meta-Regression 54
 - 6.3 Summary Meta-Regression 57

- 7 Empirical Exploration** **59**
 - 7.1 Methodological Approach Empirical Exploration 59

Contents

7.2	Execution Empirical Exploration	60
7.2.1	Participants	61
7.2.2	Data Analysis	61
7.2.3	Results	61
7.3	Summary Empirical Exploration	64
8	Discussion and Conclusion	65
8.1	Discussion	65
8.2	Conclusion	67
	Bibliography	69
A	Appendix A - Methodological Approach	77
A.1	Derivation Literature Data Evaluation	77
B	Appendix B - Quantitative studies	79
B.1	Overview Quantitative Studies Meta-Analysis	80
C	Appendix C - Results Meta-Analysis	85
C.1	Grubbs' test	85
C.2	Results evaluation criterion "time"	87
C.3	Results evaluation criterion "errors"	88
C.4	Results evaluation criterion "cognitive load"	89
C.5	Publication Bias	90
C.5.1	Duval and Tweedie's Trim and Fill	90
C.5.2	Egger's Regression of the Intercept	90
D	Appendix D - Results Meta-Regression	91
E	Appendix D - Empirical Exploration	93
E.1	Results Empirical Exploration	93

List of Figures

1.1	Tensions in the market of manufacturing companies following Teubner et al. [2018]	1
1.2	Structure of the report	3
2.1	Types of information following Feldmann et al. [2007, p. 21]	6
2.2	' <i>Virtuality continuum</i> ' following Milgram and Kishino [1994]	8
2.3	Empowerment of flexibility by AR systems [Vernim et al., 2016, p. 570]	9
2.4	Percentage of studies utilising certain measures [Egger and Masood, 2020, p. 10]	12
2.5	Frequency of subjective measures of cognitive load across articles [Jeffri and Awang Rambli, 2021, p. 6]	12
2.6	Overview existing literature reviews	13
4.1	Methodological approach	19
4.2	Framework systematic literature review by Vom Brocke et al. [2009]	20
5.1	Methodological approach meta-analysis	26
5.2	Taxonomy following Cooper [1988]	26
5.3	STARLITE approach following Booth [2006]	27
5.4	Evaluation of data following Vom Brocke et al. [2009]	28
5.5	Exemplary funnel plot no publication bias (top) and publication bias (bottom) [Sutton et al., 2000, p. 1575]	32
5.6	Exemplary trim-and-fill method including missing studies [Hopkins and Smail, 1999, p. 1575]	32
5.7	Completed taxonomy following Cooper [1988]	34
5.8	Keywords and search string combinations	35
5.9	Completed STARLITE approach following Booth [2006]	35
5.10	Resulting literature per evaluation level	36
5.11	Country distribution final paper set	37
5.12	Distribution of years final paper set	38
5.13	Distribution classical instructions	38
5.14	Distribution AR supported instructions	38
5.15	Distribution domains of relevant studies	39
5.16	Distribution experimental settings	39
5.17	Distribution of evaluation criteria	39
5.18	Influence Z- and <i>p</i> -values on null hypothesis [Sekaran and Bougie, 2016, p. 382] .	44
5.19	Forest plot for 'time'	45
5.20	Forest plot for 'errors'	45
5.21	Forest plot for 'cognitive load'	45
5.22	Funnel plot for 'time' (random effects model)	49
5.23	Trim-and-fill for 'time' (random effects model)	49
5.24	Funnel plot for 'errors' (random effects model)	50
5.25	Trim-and-fill for 'errors' (random effects model)	50
5.26	Funnel plot for 'cognitive load' (random effects model)	51

List of Figures

5.27	Trim-and-fill for 'cognitive load' (random effects model)	51
6.1	Relationship cognitive load, performance, and errors following Abbas et al. [2020]; Lyell et al. [2018]; Lindblom and Thorvald [2014]	55
6.2	Regression 'cognitive load' on 'time'	55
7.1	Survey approach following the Delphi method	60
A.1	Derivation evaluation of data following Vom Brocke et al. [2009]	77
C.1	Grubbs' test for evaluation criteria "Time", "Errors", and "Cognitive load"	85
C.2	Grubbs' test for evaluation criteria "Errors"	86
C.3	Results random effects model for evaluation criteria "time" (extracted from CMA)	87
C.4	Results one study removed for evaluation criteria "time" (extracted from CMA)	87
C.5	Results random effects model for evaluation criteria "errors" (extracted from CMA)	88
C.6	Results one study removed for evaluation criteria "time" (extracted from CMA)	88
C.7	Results random effects model for evaluation criteria "errors" (extracted from CMA)	89
C.8	Results one study removed for evaluation criteria "cognitive load" (extracted from CMA)	89
D.1	Regression country on "time"	91
D.2	Regression year on "errors"	91

List of Tables

4.1	Methods, objectives, and expected findings	21
5.1	Overview final paper set	40
5.2	Overview Grubbs' test for outliers in final paper set	42
5.3	Statistics meta-analysis	46
5.4	Heterogeneity statistics for each evaluation criteria ($p < 0.001$ for Q)	49
5.5	Summary meta-analysis	52
6.1	Regression statistics evaluation criteria	55
6.2	Regression statistics study characteristics	57
6.3	R^2 values of statistically significant moderator variables for $p < 0.1$	58
7.1	Results empirical exploration <i>potentials</i> and <i>challenges</i>	63
7.2	Results empirical exploration <i>influencing variables</i> and <i>use cases</i>	63
8.1	Results of hypotheses	68
B.1	Overview final paper set	80
B.2	Overview data final paper set	82
C.1	Results Duval and Tweedie's Trim and Fill (extracted from CMA)	90
C.2	Results Egger's Regression of the Intercept (extracted from CMA)	90
E.1	Results individual rankings question related to potentials of AR	93
E.2	Results individual rankings question related to challenges of AR	94
E.3	Results individual rankings question related to influencing variables of AR	95
E.4	Results individual rankings question related to use cases of AR	96

Acronyms

AR Augmented Reality v

MR Mixed Reality 7

HMD Head-Mounted Displays 8

NASA-TLX NASA Task Load Index 11

NASA-RTLX NASA Raw Task Load Index 11

OEE Overall Equipment Effectiveness 11

CMA Comprehensive Meta-Analysis 21

EEG Electroencephalography 67

1 Introduction

This thesis is motivated by the desire to gain a better scientific understanding of a practical problem. As a starting point, Section 1.1 describes the underlying motivation. In particular, this section highlights the tensions in the market of manufacturing companies and strengthens the usage of AR as a potential solution to those tensions. Following, the research objective and the research questions are described in Section 1.2. Lastly, Section 1.3 indicates the structure of this thesis.

1.1 Motivation

Manufacturing companies are confronted with increasing variants and individualised products, with high-quality requirements and short product life cycles [Lušić et al., 2016]. These companies find themselves in a field of tension between multiple requirements from the buyers' market and the labour market [Teubner et al., 2018]. Figure 1.1 displays the described tension in the market of manufacturing companies.

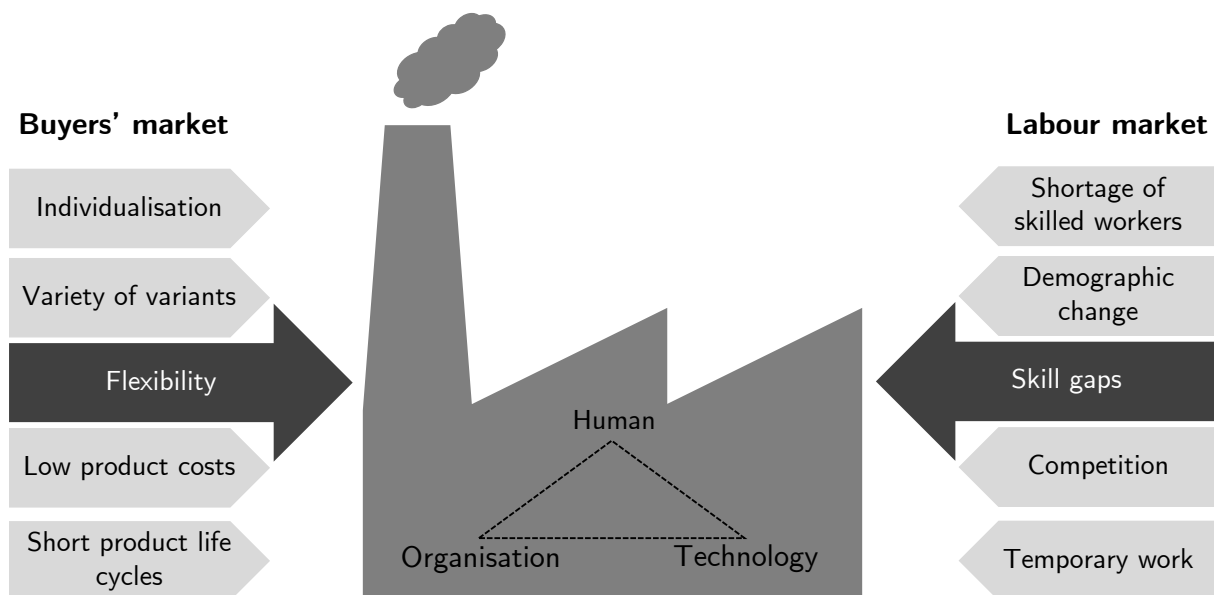


Figure 1.1: Tensions in the market of manufacturing companies following Teubner et al. [2018]

The heightened product diversity leads to interrupted learning curves, especially in maintenance applications, assembly, and machinery repair as part of manufacturing processes [Gaimon and Singhal, 1992; Masoni et al., 2017]. The management of process complexity is further challenged by an ageing and heterogeneous workforce [Hold et al., 2017]. Despite these growing challenges, manufacturing systems must be reconfigurable and flexible to react quickly to changes in the buyers' market [ElMaraghy et al., 2013].

1 Introduction

Highly experienced operators often meet the demand for flexibility with programming, maintenance, and diagnostic skills [Sethi and Sethi, 1990]. Human beings are still indispensable due to their cognitive abilities and flexibility. In particular, experienced operators can achieve flexible adaptation to changing situations and requirements. This ability to change can hardly be realised economically and technically by automated solutions [Stoessel et al., 2008].

Simultaneously, operators are exposed to alleviated increasing cognitive and psychological load due to highly flexible employee deployment and continually changing working environments and methods [Vernim and Reinhart, 2016]. The underlying information processes must be optimised to reduce both mental and psychological load in a manufacturing environment [Chen et al., 2017]. However, many manufacturing companies still find themselves confronted with an impractical and inefficient presentation of information on the shop floor [Burggräf et al., 2020]. To overcome these challenges, Industry 4.0 solutions that support employees in an agile production environment are promising [Johansson et al., 2018]. In particular, AR applications offer a way to support the interaction between people and technology and combine the advantages of manual and automated processes [Burggräf et al., 2020].

Cognitive worker assistance systems, including AR solutions, offer the potential to increase manufacturing systems' productivity and agility [Keller et al., 2019]. These devices enable efficient information distribution and support employees in the perception, reception, and processing of information [Syberfeldt et al., 2017]. In this context, the individual roles of employees, their qualifications, and personal characteristics are decisive. Taking them into account enables the provision of specific information adapted to the user and the environment [Galaske and Anderl, 2016]. In this way, an optimal distribution of information on the shop floor can be realised, strengthening manufacturing processes' competitiveness in high-wage geographical locations [Dachs et al., 2019].

1.2 Research Objective and Questions

Individual studies in the scientific literature often describe the positive effects of AR. Yet, little is known about the actual impact on employees' performance or cognitive load levels in manufacturing environments. Therefore, this research aims to explore the interrelationships in the usage of AR solutions in manufacturing environments.

First, characteristics of AR solutions in manufacturing environments will be identified to allow a quantification of the impact on variables relevant to manufacturing processes. Following, the target variables can be linked to AR solutions' characteristics.

Given the motivation and the research objective of this study, the following central research question for this thesis arises:

Can the use of AR solutions benefit manufacturing activities and if so, how?

Based on the central research question, further sub-research questions can be derived to be able to answer the central research question:

- 1. Which factors in manufacturing activities can be influenced by the use of AR solutions?*
- 2. Can those factors be measured and if so, how?*
- 3. Can a benefit be achieved and if so, by how much?*

1.3 Report Structure

This thesis is divided into eight chapters that follow the structure shown in Figure 1.2. In this context, Chapter 1 introduces this thesis and comprises three independent sections. In particular, Section 1.1 highlights the underlying motivation. Then, Section 1.2 formulates the research objective and the corresponding central research question. Three sub-questions support the central research question. Furthermore, this section presents the structure of the report.

Chapter 2 presents the theoretical foundations relevant to this thesis. Especially, the relevance of information in manufacturing environments and different fields of application of AR in such environments are described in Section 2.1 and 2.2. Subsequently, Section 2.4 identifies the knowledge gap following the existing approaches presented in Section 2.3.

Following the insights gained in the literature review, Chapter 3 describes the development of the hypotheses. First, Section 3.1 highlights the requirements for a hypothesis. Then, Section 3.2 formulates multiple hypotheses based on further insights from the literature.

Thereupon, Chapter 4 explains the methodological approach of this thesis. The methodological approach forms the foundation to answer the central research question and to prove the hypotheses.



Figure 1.2: Structure of the report

Next, the meta-analyses in Chapter 5 form the main part of this thesis. The first part outlines the scientific background of a meta-analysis. Particularly, Section 5.1 discusses the advantages and weaknesses in addition to the methodological approach and requirements. In the second

1 Introduction

part, Section 5.2 executes the meta-analyses to provide an empirical answer to the central research question. For this purpose, a systematic literature review is first carried out following an evaluation scheme to identify relevant studies. Subsequently, the studies and the corresponding data are analysed in Section 5.2.4. Finally, Section 5.5 concludes this chapter with a summary of the results of the meta-analyses.

Based on the results of the meta-analyses, Chapter 6 shows multiple meta-regressions. First, Section 6.1 summarises the theoretical background and the variables of interest. Then, Section 6.2 presents the results of the meta-regressions.

The findings from the meta-analyses and meta-regressions are then put into an industrial context with the help of expert interviews. Chapter 7 is divided into the methodological approach (Section 7.1) and the execution and evaluation of the expert interviews (Section 7.2).

Lastly, Chapter 8 concludes this thesis. The results of the report are summarised and critically reflected in Section 8.1. Finally, Section 8.2 provides a conclusion by referring back to the hypotheses raised.

2 Literature Review

Companies across industries counteract increasing competition with increasing variants and individualised products, with high-quality standards and shorter product cycles. As a consequence, the expectations and requirements in manufacturing processes increase for employees. AR applications are promising solutions to create a more productive working environment. In the first place, it is crucial to understand the challenges and variables relevant in manufacturing environments. This includes the relevance of information in such environments. Based on a comprehensive understanding of manufacturing environments and the relevance of information, emerging AR systems can be implemented in different fields of application.

Following, this chapter first presents theoretical knowledge about manufacturing environments including the relevance of information in Section 2.1.1. Thereupon, Section 2.2 includes a short definition of AR and distinguishes between different fields of application of AR. Then, Section 2.3 focuses on existing approaches addressing the influence of AR solutions in manufacturing environments. Lastly, the underlying knowledge gap is highlighted in Section 2.4.

2.1 Manufacturing Environments

Manufacturing environments encompass all value-creating and supporting processes in which a given material is transformed into a product of various shapes and sizes [Kaushish, 2010, p. 3]. Both the value-creating and supporting processes can include maintenance, production, quality, logistics, and assembly tasks and can be supported by a variety of tools, equipment, and human effort [Kaushish, 2010, p. 3].

As indicated, manufacturing companies face constantly growing competition in highly turbulent and volatile markets [Gröger et al., 2013, p. 205]. To overcome those challenges in an ever more globalized world, manufacturing companies develop more and more towards customised products. As a result, better responsiveness and flexibility is required of manufacturing companies and its employees [Spath et al., 2013, p. 42]. Accordingly, the human effort is of essential importance in manufacturing environments although capacity is limited [Stoessel et al., 2008, p. 245].

Kaushish [2010, p. 3] defines four manufacturing attributes that are considered during most human-driven decisions in manufacturing environments. These include cost, time, quality, and flexibility. An efficient supply of information supports all four attributes as shown in the following section.

2.1.1 Relevance of Information in Manufacturing Environments

Information are valuable commodities and are becoming increasingly important in manufacturing environments. In some cases, having or not having information can even be a success-critical and competition-determining criterion [Willke, 2001]. In particular, information can

2 Literature Review

result in the following effects: Immateriality as information can be used indefinitely [Wright, 1976, p. 298], the possibility of parallel usage leading to synergy effects [Mohr, 1997, p. 14], and low reproduction costs leading to potential economic advantages [Krcmar, 2015, p. 16].

Feldmann et al. [2007, p. 21] distinguish between six different types of information in industrial contexts (see Figure 2.1). These include general, procedure related, process related, order related, product related, and quality related information. The individual types and characteristics are described in the following. In practice, different kinds of information can occur at the same time.

General information. General information relate to all information that is not directly related to processes, procedures, products, or orders [Feldmann et al., 2007, p. 22]. These include general cleaning guidelines and shift schedules.

Procedure related information. Procedure related information support employees to follow standardised procedures which cover the order in which employees, resources, and material pass through manufacturing environments [Feldmann et al., 2007, p. 22].

Process related information. Process related information closely relate to procedure related information. Process related information guide the workers during their activities [Feldmann et al., 2007, p. 21f.]. As an example, these information include work instructions to operate equipment.

Order related information. Among other data, each order includes information about the products to be manufactured, number of pieces, recipient, and deadlines. The necessary data to fulfill an order is grouped as order related information [Feldmann et al., 2007, p. 21].

Product related information. Product related information contain all information that directly describe the characteristics of a product and can be assigned to it. Data sheets and operating instructions, for example, can be mentioned at this point. Order related and product related information can overlap in many cases.

Quality related information. In order to be able to guarantee the highest possible production quality, specific quality specifications must be communicated and adhered to [Feldmann et al., 2007, p. 22]. These include, for example, audit instructions or packaging regulations.

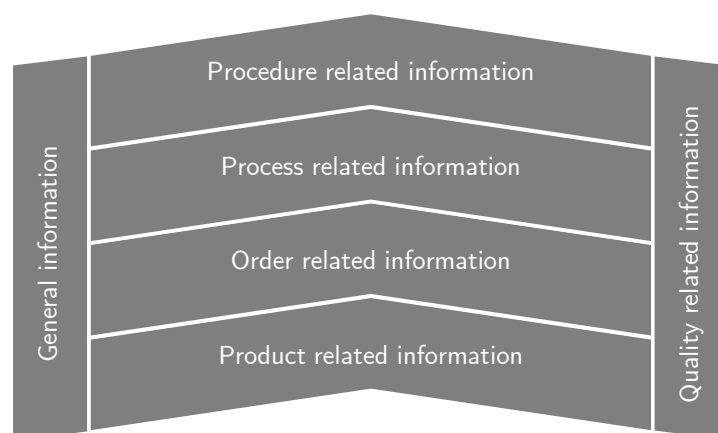


Figure 2.1: Types of information following Feldmann et al. [2007, p. 21]

Efficient processing of all types of information by individual employees highly depends on the supply of information. Feldmann et al. [2007, p. 25f.] identify seven basic requirements that support an efficient distribution of information in manufacturing environments. To allow an efficient supply of information, information must be correct, complete, punctual, comprehensible, archivable, ergonomic, and up-to-date [Feldmann et al., 2007, p. 25f.]. Individual problems can occur if basic requirements cannot be met. Hinrichsen and Bendzioch [2019, p. 341] highlight five problems of information presentation:

1. Lack of information within the work system
2. Irrelevance of a subset of the information provided in the work system
3. No process orientation: Presentation of information does not match process sequences
4. Outdated information
5. No compatibility of information representation with human modes of information processing

In practice, paper-based information is still most common, although these lack dynamic adaptation to specific needs and are highly inflexible [Feldmann et al., 2007, p. 20]. As a result, 'uncoordinated waiting times, long decision-making processes, inflexibility, and costly communication' [Gröger et al., 2013, p. 205] can occur on the shop floor. AR technologies are promising solutions to overcome barriers as part of the presentation of relevant information.

2.2 Augmented Reality in Manufacturing Environments

This section aims to introduce the technology of AR in manufacturing environments. To achieve this purpose, the technology is first defined in Section 2.2.1. Thereupon, Section 2.2.2 highlights different fields of application of AR in manufacturing environments.

2.2.1 Definition Augmented Reality

Boeing employees Tom Caudell and David Mizell first introduced the term AR in 1992 [Mizell, 2020]. However, until 1994 no uniform naming for reality-enhancing technologies nor a clear classification has been developed. In that same year, Milgram and Kishino [1994] established the concept of the *virtuality continuum* which is still valid today. The concept of a *virtuality continuum* describes the connection of completely real environments with completely virtual ones (see Figure 2.2). The focus of this project falls into the *virtuality continuum* and lies on AR respectively Mixed Reality (MR). According to Azuma [1997, p. 356], both AR and MR fulfill the following three basic characteristics:

1. Reality is combined with virtual objects and information
2. Combination takes place in real time and is interactive
3. Inserted objects are registered three-dimensional

Even though both terms generally refer to the expansion of all human senses, the current understanding is mainly limited to visual perception. The visual perception allows to display all six types of information identified by Feldmann et al. [2007, p. 21] to support the users. In practice, Head-Mounted Displays (HMD), tablets, smartphones, and stationary displays enable the access of all types of information. Depending on the enabling device, the interaction between the users and the technologies is supported by voice-control, eye-control, and head movement-control [Danielsson et al., 2020, p. 1299]. In manufacturing environments, the central objective of AR enablers is to minimize the discrepancy between human performance and the requirements of a work task and to enable flexible work deployment across different workplaces and production lines.

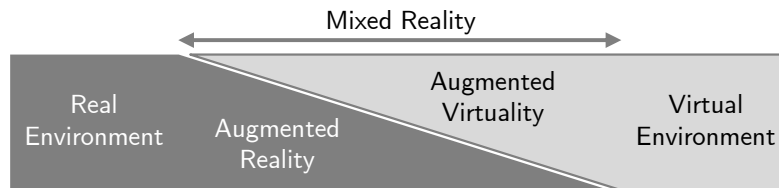


Figure 2.2: 'Virtuality continuum' following Milgram and Kishino [1994]

2.2.2 Fields of Application Augmented Reality in Manufacturing Environments

The interest in AR in manufacturing environments has increased considerably. Among other experts, Dan Arczynski predicts 'measurable savings [...] because [AR] is speeding up processes' [David Greenfield, 2017]. In theory, there are numerous use cases of AR to achieve these savings in manufacturing processes. Following, this section identifies potential fields of application.

Multiple fields of application of AR in manufacturing can be identified in the literature. In particular, Kohn and Harborth [2018] and Egger and Masood [2020] follow a traditional product manufacturing life cycle including the following phases: Planning and design, prototyping, maintenance and inspection, assembly, quality, and logistics.

Planning, design, and prototyping. The first step of the life cycle is planning and design. This phase includes the engineering of the products and the design of corresponding documents [Kohn and Harborth, 2018]. Next, prototype building and testing are identified as further fields of application. Although both phases can be supported by visual overlays with AR enablers, these fields of application still lack efficient solutions [Kohn and Harborth, 2018].

Training. The demographic change in manufacturing companies, but especially increasingly complex tasks, require efficient training. AR solutions can be a powerful tool to support complex training. Employees are enabled to directly link instructions to the tasks to strengthen and learn new skills [Webel et al., 2013].

Maintenance and inspection. Kohn and Harborth [2018] and Egger and Masood [2020] identify maintenance and inspections as a popular field of application of AR solutions. Typically, remote communication and assistance and the visualization of former paper-based instructions represent suitable applications Plakas et al. [2020]. As identified by [Kohn and Harborth, 2018], most industrial projects in Germany focus on this field of application.

Assembly. In contrast, assembly tasks represent the most popular use case for scientific research projects. The difficulty of assembly tasks continually increases due to the growing complexity in products. In this field of application, suggestions of corrective measures can

significantly improve error rates once a mistake is detected. Thus, the impact of AR is assessed to be highest in assembly [Plakas et al., 2020]. However, such statement must be proven to be statistically significant.

Quality and logistics. Lastly, Egger and Masood [2020] identifies product control and distribution as further fields of application. However, logistics, is often excluded from the manufacturing life cycle as it is an extensive research area for itself.

The described fields of application are all subject to increased requirements as a result of increasing variants and individualised products. In this context, Vernim et al. [2016, p. 570] identify assistance systems including AR systems as a key driver for flexibility (see Figure 2.3). In the context of Industry 4.0, AR systems enable the connection of smart objects and workers. AR systems enable both a new way of learning and the decentralisation of productions systems and data. As described earlier, the visualisation of necessary information is crucial to increase flexibility among different fields of application [Vernim et al., 2016, p. 570].

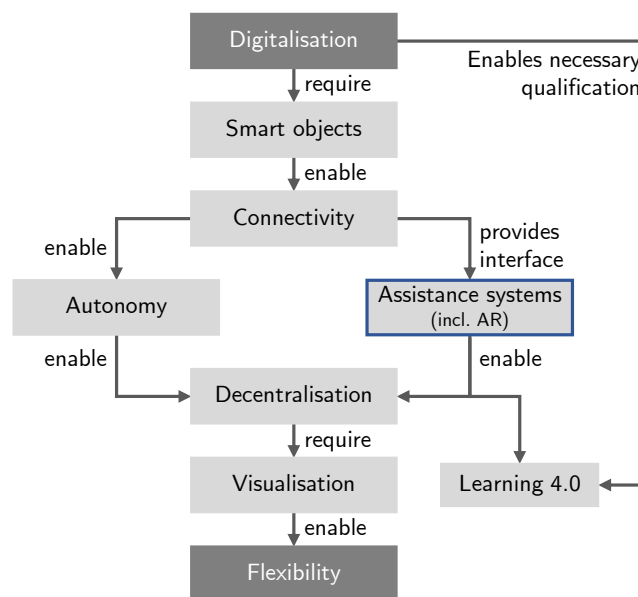


Figure 2.3: Empowerment of flexibility by AR systems [Vernim et al., 2016, p. 570]

2.3 Existing Approaches

Recently published independent systematic literature reviews by Egger and Masood [2020], Kohn and Harborth [2018], Baroroh et al. [2020], and Jeffri and Awang Rambli [2021] form the basis for this chapter. The underlying reviews all address the influence of AR solutions in manufacturing environments. However, the focus and underlying references differ in most parts. The reviews provide an holistic overview of past research activities including empirical studies in the field of AR to answer specific research questions across industries and types of technologies.

The papers highlight the results of individual user surveys addressing the current knowledge, future challenges, and influence of AR. Following Egger and Masood [2020], particularly Germany, Italy, and Singapore are geographical regions in which AR systems are being researched a lot. The majority of the research has been undertaken in laboratory environments and show

2 Literature Review

different results depending on the experimental settings. Danielsson et al. [2020] and Kohn and Harborth [2018] support the argument that current AR systems still lack standardisation and development. Even in experimental settings, different devices and use cases can change the significance of individual experiments.

As indicated, the chosen scientific articles include or address empirical studies on AR's influence in manufacturing environments. However, the reviews still lack statistical evidence as empirical results of individual user studies are not yet combined. In particular, Egger and Masood [2020] and Kohn and Harborth [2018] reveal the need to obtain a more powerful understanding of the impact of AR in manufacturing environments as some empirical studies result in ambiguous findings.

2.3.1 Relevant Factors and Measurement

The usage of AR solutions can influence different factors in manufacturing activities. To be able to make a statistically significant statement about the benefit of AR solutions, those influenced factors must first be identified as stated in the first sub-research question. Next, a way to measure the relevant factors while guaranteeing statistical comparability can be determined. As part of their paper, Egger and Masood [2020] have intensively investigated studies on AR applications in manufacturing environments.

Egger and Masood [2020] show the percentage of papers addressing a certain dependent variable. As shown in Figure 2.4, time, error rates, and cognitive load represent the most widely spread variables. Alongside these three variables, user surveys, marker decoding distance and time, head movements, and welt location can play a role. The meta-analyses focus on the top three relevant variables to find sufficient empirical studies to run the analyses. Most studies do not purely focus on the assessment of only one measure but use more than one.

The following paragraphs describe the variables processing time, error rate, and cognitive load in the context of manufacturing environments. Furthermore, the way how the respective variables can be measured is described and, thus, the second sub-research question can be answered.

Processing time. The processing time relates to the time it takes to process a task [Gong et al., 2018]. In manufacturing environments, these tasks include assembly, maintenance, and changeovers tasks. As indicated in Figure 2.4, the processing time is the most prominent measure used in past studies. This is strengthened by the fact that according to Kohn and Harborth [2018], Vanneste et al. [2020], and Baroroh et al. [2020], most studies focus on the performance improvements of individual tasks through the usage of AR systems. Time improvements directly relate to performance improvements in the context of manufacturing activities.

The processing time is measured in seconds, minutes, or hours. Consequently, comparability between different ways of information distribution for individual tasks is given. In many companies, the times are documented manually or digitally by the workers or calculated using the total number and total time.

Error rate. In the context of AR systems in manufacturing environments, error rates refer to the percentage of errors made in relation to the entirety [Kohn and Harborth, 2018, p. 7]. The focus of this research lies on human-caused errors in the context of human-machine interaction. These kinds of errors often occur due to a lack of information, knowledge, or competence.

Manufacturing companies more and more follow a zero-defect strategy as troubleshooting can be particularly costly in complex value chains.

As indicated, the error rate is usually referred to as a relative number. However, depending on the experiment, the error rate can be an absolute number as well. In practice, both scales allow comparability of the improvements of an AR system. The error rates are usually determined by counting.

According to Egger and Masood [2020] and Jeffri and Awang Rambli [2021], most studies focus on the improvements of individual tasks through the usage of AR systems. Both the processing time and the error rate relate to the performance of functions and directly influence the Overall Equipment Effectiveness (OEE) in the manufacturing area. The OEE supports the identification of losses, bench-marking of processes, and the improvements of productivity.

Cognitive load. Following Sweller [1988], cognitive load refers to the used amount of working memory resources during a task. The working memory is responsible for problem-solving and information processing and is a crucial resource in flexible manufacturing environments. Working memory capacity is limited, and information can only be maintained and processed up to a certain degree [Sweller, 1988]. Thus, the aim of instructions and AR systems should be to reduce the amount of working memory captured.

Jeffri and Awang Rambli [2021] focus on identifying relationships between the effects of AR in the context of cognitive load and task performance. To gain a more powerful understanding of AR systems, Jeffri and Awang Rambli [2021] reviewed 64 articles investigating the effects of the use of AR systems. The first research question of Jeffri and Awang Rambli [2021] relates to the methods of how cognitive load can be measured in studies evaluating the use of an AR system. Although cognitive load measurements are not directly possible, different methodologies to approximate it are established in the literature. Jeffri and Awang Rambli [2021] show that 30 out of 40 experiments make use of the NASA Task Load Index (NASA-TLX) or NASA Raw Task Load Index (NASA-RTLX) to measure cognitive load levels (see Figure 2.5).

Both the NASA-TLX and the NASA-RTLX are well-accepted frameworks in the literature to approximate cognitive load levels [Hart and Staveland, 1988, p. 6]. Both frameworks are subjective assessment tools to perform subjective workload assessments during human-machine interactions [So, 2020]. The score has a rating scale from zero to one hundred and is easily comparable [Hart and Staveland, 1988, p. 6]. The score is determined by answering a questionnaire.

2.4 Knowledge Gap

The approaches described in the previous section illustrate the relevance and effects of AR systems in industrial practice. Existing literature reveals the diversity of possible applications of such systems which results from the requirements of the respective use cases. Figure 2.6 provides an overview of the reviews presented in Section 2.3. The overview visualises the focus of the reviews, the existence of a theoretical and empirical aggregation, and the relevant variables according to the respective authors.

As described earlier, the effects of AR solutions are justified theoretically in the literature but are not yet substantiated by aggregated empirical results. Simultaneously, Terhoeven et al. [2018], Vanneste et al. [2020], and many more show individual small user studies in which the influence of AR solutions on different use cases in manufacturing environments is assessed empirically. Among other factors, the influence is particularly tested with regard to processing

2 Literature Review

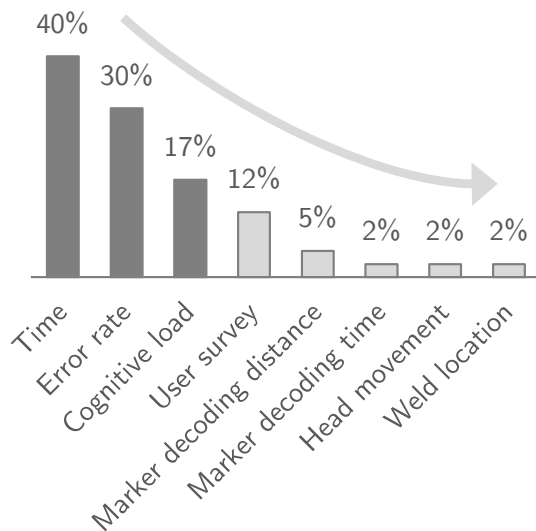


Figure 2.4: Percentage of studies utilising certain measures [Egger and Masood, 2020, p. 10]

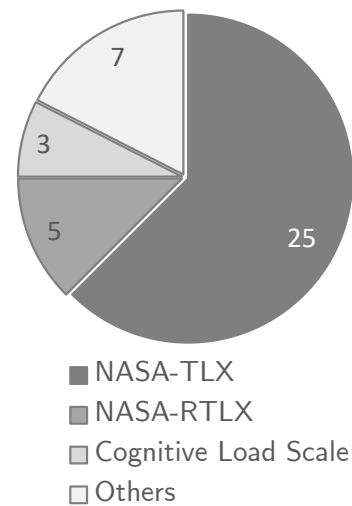































Figure 2.5: Frequency of subjective measures of cognitive load across articles [Jeffri and Awang Rambli, 2021, p. 6]

times, error tolerance, and cognitive load Egger and Masood [2020, p. 10]. As a matter of fact, sample sizes differ and empirical results show great differences depending on the principle and the use case.

A common feature underlying all experiments is that they have not yet been investigated in practice-relevant, long-term field experiments. Furthermore, individual studies show ambiguous results, and a statistically powerful empirical assessment is still missing. The aggregated assessment of empirical studies allows a statistical comparison of surveys and experiments dealing with similar research questions. Nevertheless, the results heavily depend on the operators' cognition and experience. Regardless of the effects on individual factors, comparability of user studies during the summary of empirical studies needs to be kept in mind.

Egger and Masood [2020], Kohn and Harborth [2018], Baroroh et al. [2020], Vanneste et al. [2020], and Jeffri and Awang Rambli [2021] highlight that an efficient implementation of AR in manufacturing environments still requires additional research. In particular, a powerful aggregation of quantitative-empirical effects of such technology and possible relationships of effected variables reveal a knowledge gap (see Figure 2.6). For this reason, a meta-analysis is needed to determine the influence of AR by synthesising several small empirical studies into one large study.

	Egger and Magnoo (2020)	Kohn and Harborth (2018)	Baroroh et al. (2020)	Jeffri and Awang Rambli (2021)
 Content (domain)	SLR ¹ (manufacturing)	SLR (manufacturing)	SLR (manufacturing)	SLR (across industries)
 Technological focus	AR ²	AR	AR	AR
 Theoretical aggregation	✓	✓	✓	✓
 Empirical aggregation	✗	✗	✗	✗
 Relevant variables	  	  	  	  
	  	  	  	  

1 : Systematic Literature Review 2 : Augmented Reality ✓ Included ✗ Not included ⌚ Time ✗ Error rate 🧠 Cognitive load

Figure 2.6: Overview existing literature reviews

3 Hypothesis Development

This chapter focuses on the development of three hypotheses that will be answered as part of this research project. Prior to the derivation of the hypotheses in Section 3.2, Section 3.1 provides a brief definition of hypotheses and highlights the requirements.

3.1 Definition Hypothesis

A research hypothesis describes a proposition or predictive statement addressing a possible research outcome based on expected differences between at least two variables of interest [Allen, 2017]. The cause-effect relationships are developed taking into account existing knowledge [Basavanna, 2015, p. 32]. Depending on how the hypotheses are formulated, they can be decisive for the design of the research project. In any case, empirical studies help to test the hypotheses.

A hypothesis is subject to various requirements, the four most important of which are explained in the following. First of all, a good hypothesis 'must state an expected relationship between variables' [Allen, 2017]. The expected relationship should be formulated in a positive way, assuming that a relationship exists rather than stating that it does not. Secondly, a hypothesis must be testable and falsifiable with scientific methods [Allen, 2017]. Third, a good hypothesis must be logical and based on previous theories or observations [Basavanna, 2015, p. 34f.]. In line, a logical chain of reasoning is important. Finally, in terms of language a hypothesis should be formulated in a simple and concise way to allow an easy understanding of the content [Allen, 2017; Basavanna, 2015]. Here, precision of expression should be preferred over stylistic variations [Töpfer, 2010, p. 151].

3.2 Derivation of Hypothesis

AR applications can have a significant influence on multiple measures in a manufacturing environment. As indicated, the emphasis in scientific studies often lies on the evaluation criteria 'time', 'errors', and 'cognitive load'. In the following, a hypothesis is derived for each of these evaluation criteria. The hypotheses are driven by the results in Chapter 2 and by theoretical guidance and prior evidence.

As shown in Section 2.1.1, empowering operators to absorb and process different types of information can be crucial to the success of a company. In particular, Hinrichsen and Bendzioch [2019, p. 341] list five problems of information presentation that may cause performance losses and increase cognitive load levels. Following these problems, it is indisputable to provide the right and latest information, at the right time, in the right quantity, and in the right way. Depending on how well the individual components are fulfilled, a receiver has to provide more or less capacity for absorbing and processing the information.

3 Hypothesis Development

Among others, Kock [2005] and Daft and Lengel [1986] indicate that humans are endowed with capacity-limited perceptual processes. In particular, as a result of history, individuals are limited in terms of their ability to absorb and process information. More importantly, the ability to absorb and process information highly depends on the respective information medium [Kock, 2005; Daft and Lengel, 1986]. Kock [2005] illustrates this dependence with the help of the media naturalness hypothesis.

The media naturalness hypothesis builds on the fact that humans relied almost exclusively on face-to-face communication in the past and can best process this information medium. Over time, humans have evolved cognitive, physiological, and genetic skills to process information from channels with similar characteristics as face-to-face communication. According to Kock [2005, p. 124] and Daft and Lengel [1986], these characteristics can be divided into the following five elements: Co-location (a condition in which communication partners are physically located right next to each other), synchronicity (receiving and sending information without any latency), the ability to convey facial expressions, body language, and speech.

The more elements a certain way of communication incorporates, the higher the degree of naturalness [Kock, 2005, p. 124]. Likewise, the higher the degree of naturalness, the lower the level of *cognitive effort*, *communication ambiguity*, and the higher the level of *physiological arousal* as key constructs [Kock, 2005, p. 121]. In the following, all three hypotheses are derived based on these three key constructs affected by the media naturalness hypothesis by Kock [2005].

In contrast to classical instructions (e.g., paper, PDF tablet), newly adapted AR solutions correspond to a greater extent to the described characteristics of face-to-face communication. First of all, both classical instructions and AR applications do not allow co-location as communication partners are not physically present. However, perceived virtual co-location can be achieved to a certain extent by using AR solutions [Lukosch et al., 2015, p. 519]. What is more, AR applications, in contrast to classical instructions, allow synchronicity [Liang and Roast, 2014, p. 609f.]. Users are only able to receive and send information using AR and dynamic media. Next, classical instructions do not enable users to convey facial expressions, body language, and speech during information sharing. In contrast, AR offers a wide range of possibilities to use these elements depending on the use cases [Chen et al., 2015]. Given these arguments, media naturalness is assessed to be higher for AR applications in comparison to classical ways of communication but not face-to-face. Based on the key constructs affected by the media naturalness hypothesis, reducing effects for the described evaluation criteria can be achieved.

Time

First, the most prominent evaluation criterion 'time' is considered. The corresponding hypothesis mainly refers to the key construct *physiological arousal* [Kock, 2005, p. 123]. Physiological arousal describes the degree of activation of the central nerve system of individuals [Iwańczuk and Guźniczka, 2015]. The Yerkes-Dodson law predicts that up to a certain extent higher physical arousal leads to performance improvements, including time reductions [Teigen, 1994]. This prediction applies to simple, well-learned, and partly unfamiliar tasks, which is assumed to be applicable for this research in manufacturing environments.

As a result of increased media naturalness for AR applications, individuals are more likely to be triggered by physiological arousal using this technology in comparison to classical media [Kock, 2005, p. 123]. In addition, AR helps to provide the right information, at the right time,

in the right quantity to a worker, thus increasing information availability and preventing unnecessary waste of time [Hicks, 2007, p. 239]. Given these arguments, processing times in manufacturing environments are expected to be lower while using AR solutions.

H1: AR solutions have a reducing effect on processing times of workers during manufacturing activities.

Errors

The second variable that plays a prominent role is the evaluation criterion 'errors'. Following Kock [2005, p. 124], *communication ambiguity* is one of the key constructs that is affected by media naturalness. Communication ambiguity refers to differing interpretations of individual events, including the perception and processing of information. Simultaneously, Karanikas et al. [2018, p. 261] indicate that error rates in manufacturing environments are driven by *communication ambiguity*. As a result of such ambiguity, humans tend to fill in knowledge gaps themselves if an information-giving stimulus is missing [Kock, 2005; Karanikas et al., 2018]. Subsequently, the number of errors and misinterpretations increases, especially in the case of knowledge gaps.

In the case of AR applications, the media naturalness of such technology is assessed to be higher compared to classical instructions. Based on the previously presented theoretical findings, the interpretation of data and information is thus expected to be higher, leading to fewer errors. Receivers using AR solutions can better understand the information that is meant to be communicated. At the same time, the *physiological arousal* described above also positively influences the task outcome quality Kock [2005, p. 124] and errors can be detected at an earlier stage by digital AR systems. Hence, the following second hypothesis arises:

H2: AR solutions have a reducing effect on error rates of workers during manufacturing activities.

Cognitive load

The third hypothesis is derived especially based on the affected key construct *cognitive effort*. Among others, Kock [2004] and Daft and Lengel [1986] show that decreases in media naturalness lead to increased cognitive load levels. As described previously, the level of media naturalness is assessed to be higher for AR applications in comparison to static instructions. Likewise, AR solutions can help to relieve the information recipient with lower processing requirements. In manufacturing environments, cognitive load levels can particularly be improved by AR supported information provision adapted to an individual operator. Furthermore, Kock [2005, p. 121] shows that more 'natural' communication media such as AR solutions make use of less 'specialized brain circuits to make up for the absence of [one of the five] elements'.

Given these explanations, the third hypothesis argues that following the media naturalness theory AR solutions have a positive effect on cognitive load levels of workers during manufacturing activities in comparison to static instructions.

H3: AR solutions have a reducing effect on cognitive load levels of workers during manufacturing activities.

4 Methodological Approach

The following chapter describes the methodological approach to answer the central research question and the corresponding sub-questions. The methodological approach follows the need for research as identified in Chapter 2 and 2.3, and includes three sequential phases: (1) Meta-analysis, (2) meta-regression, and (3) empirical exploration. Figure 4.1 displays the individual steps of this research. Following, the phases are explained in more detail.

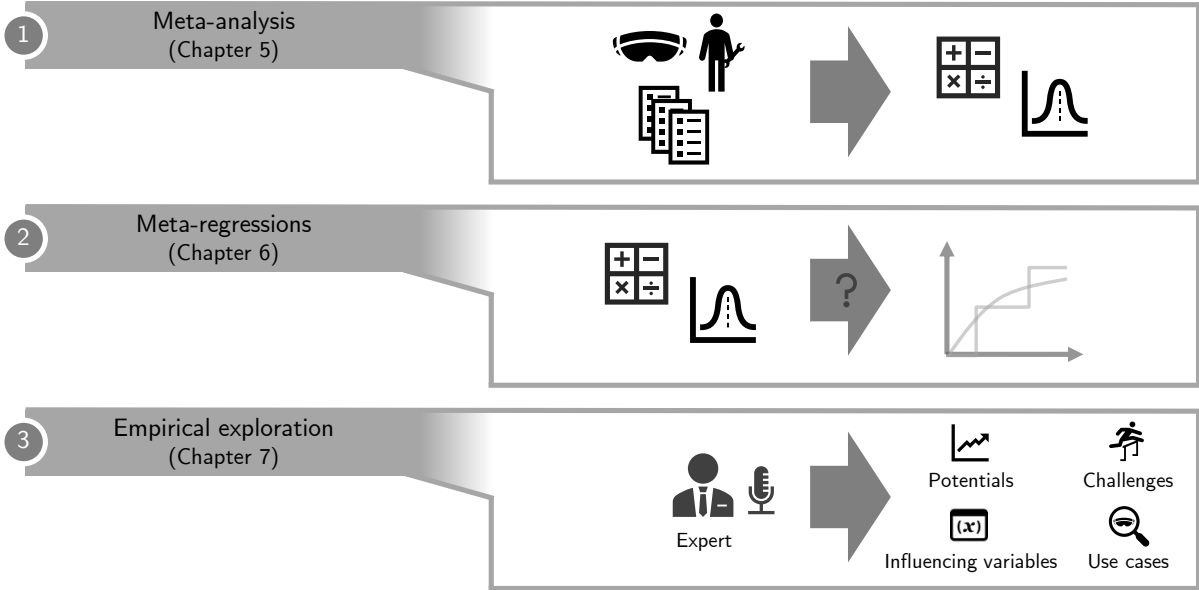


Figure 4.1: Methodological approach

1. Meta-analysis

First, three meta-analyses are carried out to analyse the state-of-the-art and the influence of AR solutions on variables relevant to manufacturing processes. As identified in Section 2.3, these variables include processing time, error rates, and the cognitive stress of employees.

A meta-analysis is used if individual studies available show ambiguous results, and a statistically powerful assessment is still missing. By synthesising several small studies into one large study, a meta-analysis provides higher significance. As a result, a more powerful statistical influence of AR on the evaluation criteria processing time, error rate, and cognitive load is expected. Meta-analyses follow six sequential phases: Formulation of the research question, data collection, evaluation of data, analysis and interpretation of data, sensitivity analysis, and presentation of results.

As part of the meta-analyses, a systematic literature search is conducted to collect and evaluate relevant data. The underlying literature search follows the framework by Vom Brocke et al.

4 Methodological Approach

[2009] (as shown in Figure 4.2), which builds on five sequential steps: The definition of the review scope, the conceptualization of the topic, the literature search, the literature analysis and synthesis, and the research agenda. Each step includes individual systematic approaches that are in line with collecting and evaluating data as part of the meta-analyses.

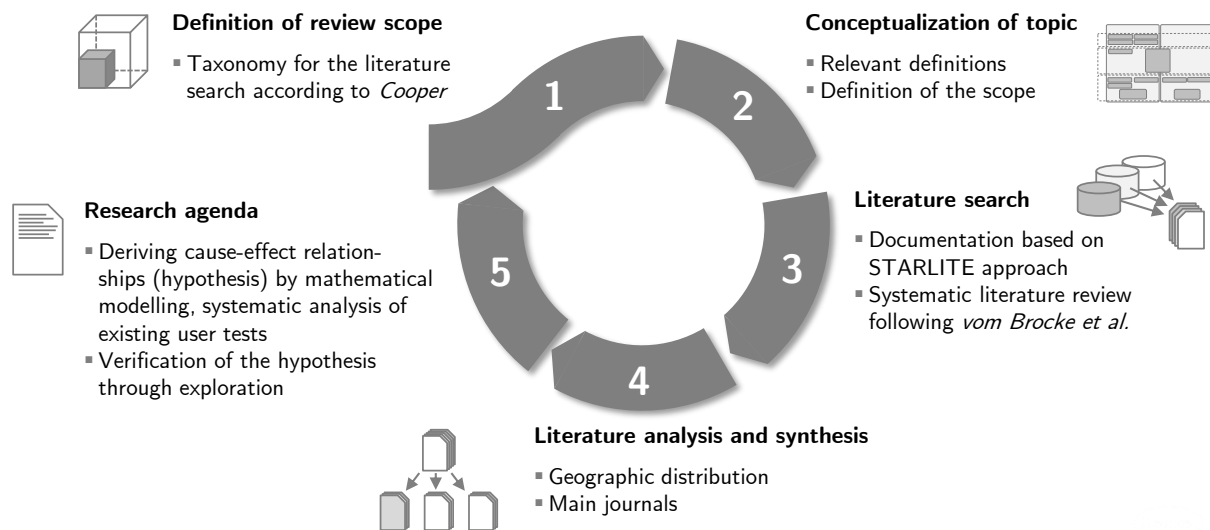


Figure 4.2: Framework systematic literature review by Vom Brocke et al. [2009]

The systematic literature review aims to build an extensive literature database covering empirical studies on AR technologies in manufacturing environments. Here, the type of technology used to enable AR is not specified in advance. Different types of technological enablers and use cases shall be compared, such that the impact and advantages of different technologies on the target variables can be distinguished. The publications contained in the database are evaluated based on an evaluation scheme (see Figure 5.4). All identified publications are classified with the help of a homogeneity assessment to avoid the 'apples and oranges problem' [Lipsey and Wilson, 2001]. The remaining studies are evaluated with regard to minimal statistical requirements to allow extraction and synthesising. Thereupon, the meta-analyses will be carried out with a previously selected software. The software helps to run statistical calculations to allow an evaluation and interpretation of the data. As part of the subsequent sensitivity analysis, the results are verified by checking for statistical heterogeneity, publication bias, and other confounding factors.

2. Meta-regression

Building on the results of the meta-analyses including the systematic literature review, multiple meta-regressions are carried out next. For each of the chosen evaluation criteria, the aim of the meta-regressions is to identify potential moderator variables to express variances between the studies included. The meta-regressions are conducted using the random effects meta-regressions model.

To explain heterogeneity, 9 meta-regressions are conducted for each of the evaluation criteria. First, the evaluation criteria 'time', 'errors', and 'cognitive load' are investigated as mutual covariates. Then, further potential moderator variables addressing the characteristics of the

studies are analysed. The meta-regressions are also carried out with the same software as the meta-analyses, namely Comprehensive Meta-Analysis (CMA).




3. Empirical exploration

Lastly, an empirical exploration will be prepared and executed based on the previously gained results. Expert surveys help to support the findings from the first two phases. The expert surveys are conducted in the manufacturing department of a chemical and consumer goods company.

The surveys follow the Delphi method, which is a systematic, multi-stage survey procedure that allows feedback from the experts. This research project identifies four areas of focus and corresponding research questions to encourage the experts to provide assessments, descriptions, and narratives on the topic.

Table 4.1 summarises the methods used, the objectives, and the expected findings of this research project.

Table 4.1: Methods, objectives, and expected findings

Method 	Objective 	Expected findings 
Meta-analysis	Analysis state of the art and AR technologies	Quantifiable influence of AR solutions on the evaluation criteria time, errors, and cognitive load
Meta-regression	Identification of linear relationships between outcome measures and covariates	Moderating variables to express variances for the chosen evaluation criteria
Empirical exploration	Validation of results in an industrial environment	Verification or adjustment of the results obtained in the meta-analyses and meta-regressions

5 Meta-Analysis

Section 2.3 identifies various scientific reviews addressing the empirical influence of AR on manufacturing environments. However, these scientific reviews lack robust statistical power. As indicated, a meta-analysis is a promising statistical tool to achieve the desired evidence. Thus, a meta-analysis can be a suitable tool to address the research gap. In the following, Section 5.1 provides a definition of meta-analysis, highlights the requirements and the methodological approach, and discusses the advantages and weaknesses. Next, three meta-analyses are carried out and described in Section 5.2.

5.1 Definition Meta-Analysis

A meta-analysis is a statistical technique for combining the results of individual scientific articles on the same topic that have been systematically researched [Jesson et al., 2011, p. 130]. In this way, a meta-analysis enables an evaluation of the overall impact of the research conducted [Pereira et al., 2019, p. 5]. The systematic approach of a meta-analysis allows for the identification of existing evidence and correlations [Jesson et al., 2011, p. 130]. For this reason, this statistical tool is an essential methodology for knowledge development and makes a valuable contribution to future research agendas [Pereira et al., 2019, p. 5].

However, whenever full data sets are available for the relevant studies, it is more suitable to analyse them directly using conventional methods rather than conducting a meta-analysis [Lipsey and Wilson, 2001, p. 2].

Differentiation meta-analysis from primary data analysis

In comparison to a primary data analysis, a meta-analysis does not require access to the raw data of a study. Instead, only the statistical values of the data from the primary study (e.g. correlation, mean) are needed. Therefore, in a meta-analysis, the number of available studies represent the unit of analysis and the statistical values correspond to the data. Individual studies show different numbers and characteristics of statistical values. Thus, mean and standard deviation are variables of interest as part of a meta-analysis [Forza and Di Nuzzo, 1998, p. 839].

5.1.1 Fields of Application Meta-Analysis

As described previously, a meta-analysis can be used to analyse statistical values from individual, but comparable studies. Following, this section identifies three different fields of application of meta-analysis: Development of theories, explanation of theories, and theoretical discussion.

Development of theories. To begin with, one application of meta-analysis is the development of theories as a result of joint research across scientific disciplines [Forza and Di Nuzzo, 1998; Pereira et al., 2019, p. 839]. In many cases, researchers draw on the results of previous studies to evaluate existing knowledge. Meta-analysis enables researchers to understand these studies better and identify knowledge gaps [Pereira et al., 2019, p. 6]. Following a meta-analysis, knowledge gaps can be further developed [Forza and Di Nuzzo, 1998, p. 839].

Explanation of theories. Secondly, meta-analysis can explain theories as well as deductive and analytical processes [Pereira et al., 2019, p. 6]. Meta-analysis can substantiate abstract study concepts by synthesising the results [Pereira et al., 2019, p. 6].

Theoretical discussion. In addition, the theoretical discussion of individual study results represents a third possible application of meta-analysis [Pereira et al., 2019, p. 6]. Here, individual studies are summarised so that the results of the meta-analysis help answer the underlying research question. The results of a meta-analysis in comparison to individual studies are more powerful [Pereira et al., 2019, p. 6].

However, different probability levels of each study must be taken into account in all three applications. According to Taveggia [1974, p. 397], empirical research is always probabilistic. Following, the results of each study might have arisen by chance. Even multiple studies may reach different conclusions due to chance variations, differences in research methods, or other errors. As a result, researchers need to determine how these 'artefacts' affect the results [Taveggia, 1974, p. 397]. These artefacts need to be removed to prevent inaccurate conclusions, and the relationships between the variables need to be identified. Then again, the process is complicated as the original data is usually not readily available for subsequent research [White, 1996, p. 325]. In general, published results from individual research studies show little more than descriptive statistics and correlation coefficients. However, meta-analysis offers various statistical options that allow an analysis of primary data studies only taking into account the published results [White, 1996, p. 325].

5.1.2 Requirements Meta-Analysis

In preparation for a meta-analysis on the influence of AR, various requirements must be kept in mind. To begin with, meta-analytical methods are only applicable to empirical research studies. Thus, performing a meta-analysis for descriptive, model-theoretical, or case-study-based research is not suitable [Lipsey and Wilson, 2001, p. 2]. Furthermore, empirical research results are required to conduct such statistical analysis. This excludes qualitative forms of research [Lipsey and Wilson, 2001, p. 2]. As a result, scientific studies using quantitative measures and presenting descriptive and inferential statistics to summarize the resulting data are suitable [Pereira et al., 2019, p. 6]. Descriptive statistics deal with collecting and observing data and their presentation in tables or graphs, for example. Inferential statistics use stochastic models to conclude the causes that caused the data described in descriptive statistics [Pereira et al., 2019, p. 6].

First, the meta-analysis requires a systematic literature review in which potential scientific studies can be identified [Jesson et al., 2011]. In this context, it is important to ensure that the selected literature is as representative as possible for the respective research area and has the necessary methodological quality [Lipsey and Wilson, 2001, p. 3]. Furthermore, the 'publication bias' must be taken into account, which is discussed in more detail in Section 5.1.5.

As described, the meta-analysis focuses on aggregating and comparing the results of individual research studies. To allow a powerful comparison, the empirical results must be

- conceptually comparable, i.e. deal with the same topics and relationships, and
- available in similar statistical forms [Lipsey and Wilson, 2001, p. 3].

However, according to Schmidt and Hunter [2016, p. 429f.], no rules are defined in the literature that detail the above-mentioned prerequisites. The research question, the research methods, and the study structure should follow similar approaches to allow comparability. In reality, the decision on comparability of individual study results is ultimately subjective and based on the opinion of the researcher [Lipsey and Wilson, 2001, p. 3]. Section 5.1.5 describes this topic in more detail.

Further requirements for the performance of a meta-analysis are based on statistical independence of variables. Independence of random variables is given as long as the occurrence of one event A has no influence on another event B. Following a strict interpretation of the independence condition, only one result from the same question should be taken from a primary study. However, researchers partly disagree on the definition of statistical independence and do not follow a strict interpretation consistently [Schmidt and Hunter, 2016, p. 429f.]. In reality, the results of various primary data studies are assumed to be statistically independent instead [Nelson and Kennedy, 2009, p. 351]. As a matter of fact, Bijmolt and Pieters [2001, p. 168] even state that a meta-analysis leads to unsatisfactory results if the statistical independence condition is strictly adhered to. Accordingly, Bijmolt and Pieters [2001, p. 168] argue against the strict interpretation of the statistical independence condition. In addition, Forza and Di Nuzzo [1998, p. 844] identify further prerequisites for conducting a meta-analysis:

- Clear and homogeneous definition of concepts
- Use of scientific measurement variables (valid, reliable, and shared)
- Provision of detailed information on sampling design and resulting samples of the primary studies
- Provision of useful information in the primary studies (such as mean and standard deviation for each variable, sample size and missing values)
- Information related to assumptions, conditions and hypotheses

What is more, the minimum number of studies needed to conduct a meta-analysis is not specified. Instead, Lipsey and Wilson [2001] argue that in some cases a meta-analysis can be applied with as few as two or three individual studies.

This section describes the necessary prerequisites for conducting a meta-analysis. Next, the methodological structure and procedure of a meta-analysis is explained in more detail.

5.1.3 Methodological Approach Meta-Analysis

As shown in Figure 5.1, the meta-analysis follows six sequential phases: (1) Formulation of a research question, (2) collection of data, (3) evaluation of data, (4) analysis and interpretation of data, (5) performance sensitivity analysis, and (6) presentation of results. In the following, this section elaborates on the individual steps in more detail.

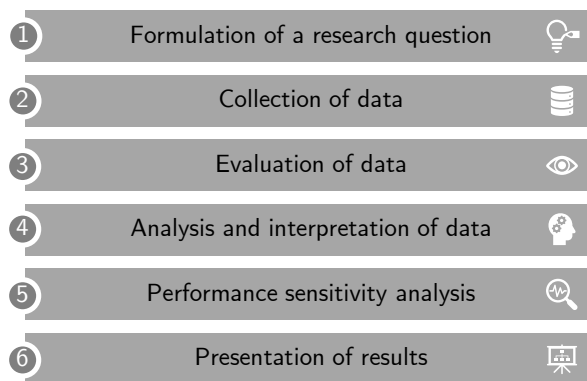


Figure 5.1: Methodological approach meta-analysis

Formulation of a research question

The first step of a meta-analysis is to formulate a research question taking into account the existing literature. As a result, only those studies that address the relevant hypothesis are considered in the following steps [Forza and Di Nuzzo, 1998; Mengist et al., 2020]. Before formulating a research question, it is helpful first to examine a variety of empirical studies [Schmidt and Hunter, 2016, p. 3]. Next, it is necessary to agree on an acceptable level of variation between individual studies [Forza and Di Nuzzo, 1998, p. 840]

Collection of data

A meta-analysis requires a structured method for selecting relevant studies and the corresponding data. In this case, the selection method is subject to special attention to avoid possible sources of bias [Forza and Di Nuzzo, 1998, p. 840]. A systematic literature search is a common approach to begin the data collection in a meta-analysis. Compared to a traditional literature search, the systematic literature search is a reproducible and transparent process and minimises such bias. It helps collect all related publications and documents that meet the predefined inclusion criteria to answer a specific research question [Mengist et al., 2020].

Cooper [1988] has proposed a taxonomy to support the classification of systematic literature searches. Figure 5.2 shows the different classification options. The taxonomy by Cooper [1988] is in line with the framework for systematic literature reviews by Vom Brocke et al. [2009] as indicated in Chapter 4.

Characteristic	Categories			
Focus	Research outcomes	Research methods	Theories	Applications
Goal	Integration		Criticism	Central Issues
Perspective	Neutral Representation		Espousal of Position	
Coverage	Exhaustive	Exhaustive and Selective	Representative	Central
Organization	Historical		Conceptual	Methodological
Audience	Specialized Scholars	General Scholars	Practitioners	General Public

Figure 5.2: Taxonomy following Cooper [1988]

Next, a search string is generated after the literature search has been classified and narrowed down with the help of the taxonomy by Cooper [1988]. The search string and various combinations of the keywords help select relevant studies that meet the selection criteria [Mengist et al., 2020, p. 2]. The STARLITE approach by Booth [2006] represents a suitable documentation standard for conducting the literature search based on predefined keywords. As shown in Figure 5.3, the scope of the literature search is defined and specified with the help of the following eight categories: Sampling strategy, type of studies, approaches, range of years, limits, inclusion and exclusion, terms used, and electronic sources.

The STARLITE methodology is widely used in scientific papers and increases acceptance of the work due to transparency. Furthermore, a structured approach reduces the risks associated with different types of bias. A systematic literature review can provide reliable findings and conclusions if the process is carried out properly [Mengist et al., 2020, p. 2]. By definition, a systematic literature review is transparent, transferable, and replicable and, for this reason, also requires a protocol. This helps to minimise distortions caused by extensive literature searches [Mengist et al., 2020, p. 3].



Figure 5.3: STARLITE approach following Booth [2006]

Evaluation of data

The evaluation of the data follows the described data collection phase and is highly dependent on the results of the latter [Forza and Di Nuzzo, 1998, p. 840]. This phase identifies suitable primary studies to be included in the meta-analysis based on the data collection phase results. The collected primary studies are assessed with the help of a predefined and systematic evaluation scheme. Consequently, individual studies are eliminated, and the relevance and statistical independence of the meta-analysis are strengthened.

As part of the framework for systematic literature reviews (Figure 4.2), Vom Brocke et al. [2009] suggest a structured literature search process. This thesis follows the given literature search process, as shown in Figure A.1 in Appendix A.1. The STARLITE methodology results form the basis of the evaluation scheme and are, thus, included in the first two steps of the scheme. Figure 5.4 displays the evaluation scheme used in preparation for the meta-analysis.

The first step of the evaluation scheme is the identification of primary studies. Primary studies are collected based on the STARLITE methodology and the corresponding keywords and search strings. Next, the duplicates are eliminated. Duplicates occur as the keyword search is conducted in multiple databases. Additionally, the results of the keyword search from off-topic journals are eliminated. Following, the eligibility of the remaining articles with regard to the

research question is evaluated. Here, the depth of content increases gradually. First, the individual titles are assessed. Second, the abstracts of the remaining primary studies are evaluated. Finally, the full-text of the remaining studies is assessed. Homogeneity and the availability of the right data are crucial at this point. Last, further primary studies are identified through a forward and backward search. The chosen studies from the forward and backward search are evaluated according to the described procedure. As a result of the evaluation scheme, a relevant and predefined literature database is created.

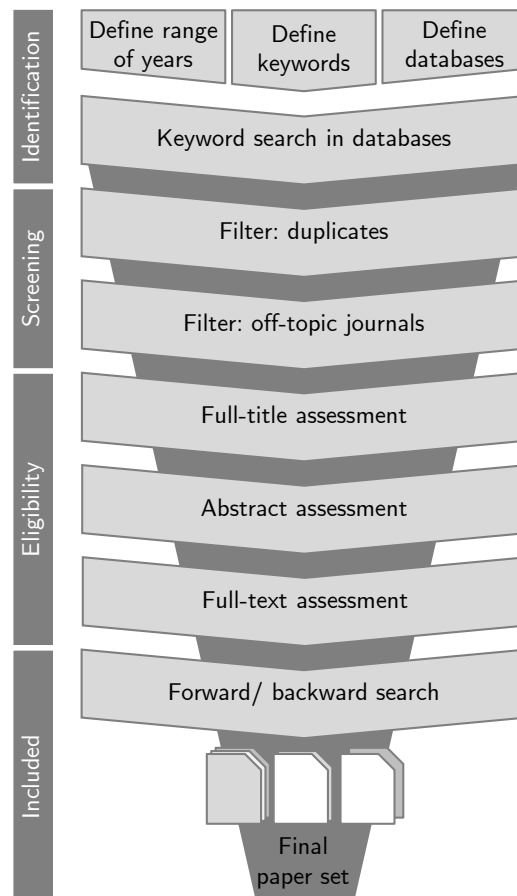


Figure 5.4: Evaluation of data following Vom Brocke et al. [2009]

Analysis and interpretation of data

Given the final paper set, the data can be analysed and interpreted. The analysis and interpretation presuppose a summary of the data. In particular, the data from selected papers needs to be extracted and standardised to derive insights and conclusions.

The data extraction process involves identifying and extracting relevant data from the selected papers [Mengist et al., 2020, p. 8]. The literature distinguishes between three models for synthesising the data: The equal effect model, the fixed effect model, and the random effect model [Forza and Di Nuzzo, 1998, p. 840f.].

The equal effect model implies that the statistical parameter of interest (e.g. correlation, mean) are the same for each study. As a result, the selected studies are controlled by the corresponding parameter [Forza and Di Nuzzo, 1998, p. 840].

The fixed effect model presupposes statistical homogeneity and assumes that the effect size of different individual studies is only different as a result of random variability. Besides that, there is only one 'true' effect for the investigated research question. This effect describes the correlations between the parameters [Forza and Di Nuzzo, 1998, p. 840].

The random effect model is used in the case that the effect sizes of individual studies vary greatly while the studies appear clinically homogeneous at the same time. Consequently, statistical heterogeneity must be assumed and tested. The tests provide information on whether the effect sizes' dispersion is greater than caused by random variability. The random effect model aims to estimate the distribution of unknown parameters [Forza and Di Nuzzo, 1998, p. 840f.].

As described, all three models calculate the weighted effect of each primary study on the overall result. The most suitable model needs to be chosen based on the data in the final paper set.

The scaling of the variables varies depending on the individual studies. Consequently, it is not possible to simply adopt the coefficient values for the analysis of the data. The data must be standardised after the extraction of the data [Schmidt and Hunter, 2016, p.193f.]. In this case, the effect size supports the statistical standardisation of empirical results. The effect size enables a standardised interpretation across all variables and measures [Lipsey and Wilson, 2001, p. 4]. Thus, the definition of suitable effect size to allow a meaningful numerical comparison and analysis across primary studies is a crucial step in meta-analysis. In general, the effect size captures the magnitude, direction, and statistical significance of the variables' relationships. The effect size should be defined so that other factors such as sample size have as little influence as possible [Lipsey and Wilson, 2001, p. 5]. The most common and suitable effect sizes in a meta-analysis are the mean and the standard deviation [Lipsey and Wilson, 2001, p. 4]. Logically, a meta-analysis can only be carried out if the respective values are available in the primary studies.

The analysis and interpretation of data includes 'the evaluation of the synthesised data and the extraction of meaningful information' Mengist et al. [2020, p. 8]. The results of the analysis can be summarised with the help of a table to increase clarity [Forza and Di Nuzzo, 1998, p. 839]. Meanwhile, at the end of this phase the research questions can be answered and explained with the help of qualitative and quantitative explanations of the results.

Performance sensitivity analysis

Following the analysis and the interpretation of the data, a sensitivity analysis enables the investigation of possible heterogeneity and publication bias [Jesson et al., 2011, p. 143f.]. The aim is to investigate whether there are significant differences between the studies and if the studies included are representative. As a result, the calculated true effect sizes of the meta-analysis can be critically assessed [Jesson et al., 2011, p. 143f.].

Presentation of results

As shown in Figure 5.1, the last step of a meta-analysis is the presentation of the generated results. The presentation of the results can take place as part of a thesis project or a scientific article [Forza and Di Nuzzo, 1998, p. 841]. The aim is to expand and complete the state of

knowledge with regard to the research question addressed and to provide a basis for future scientific research [Forza and Di Nuzzo, 1998, p. 841].

5.1.4 Advantages Meta-Analysis

The great advantage of a meta-analysis is that it is not necessary to access the raw data of the individual studies when performing the analysis. Instead, the statistical values are sufficient for analysis and allow a decisive result to be obtained [Lipsey and Wilson, 2001, p. 5f.]. However, if the complete data sets of the relevant studies are available, it generally makes more sense and is more powerful to analyze them directly with conventional methods [Lipsey and Wilson, 2001]. Still, Lipsey and Wilson [2001] highlight four advantages of a meta-analysis as described in the following.

First, a meta-analysis enables assessing the author's 'assumptions, procedures, evidence, and conclusions' [Lipsey and Wilson, 2001, p. 5f.] as the research process is explicit and systematic. A meta-analysis can provide objective and powerful results [Lipsey and Wilson, 2001, p. 5].

Second, Lipsey and Wilson [2001] emphasise that a meta-analysis is a quantitative method that presents key study findings in a more differentiated and sophisticated way compared to qualitative summaries or 'vote-counting'. By capturing the magnitude and direction of each relevant statistical relationship in primary studies, the meta-analysis allows determining the strength of the effects [Lipsey and Wilson, 2001, p. 6]. In this way, a meta-analysis allows for an analytically precise examination of the relationships between study outcomes and study characteristics.

Additionally, a meta-analysis provides higher significance by synthesising several small studies into one large study [Lipsey and Wilson, 2001, p. 6]. As a result of the synthesis, a meta-analysis helps discover meaningful effects and relationships in the literature [Lipsey and Wilson, 2001, p. 6]. This is particularly useful when an individual study's size is insufficient to show an effect or the overall effect is superficially ambiguous due to differential effects in the individual studies.

Lastly, the meta-analysis enables an organised way of handling information from a large number of primary study results [Lipsey and Wilson, 2001, p. 6]. Meta-analysis is conducted with a software program that can process almost unlimited studies with different statistical input data.

5.1.5 Weaknesses and Sources of Error Meta-Analysis

Meta-analyses are not only advantageous, but also have various weaknesses and sources of error. Possible weaknesses and sources of error occur as a result of the combination of multiple primary studies without having access to the raw data. In the following, the resulting weaknesses and sources of error are discussed.

The first weakness of meta-analyses addresses the quality of primary studies. As described previously, the results of a meta-analysis highly depend on the quality of the final paper set. As a result, methodologically poorer studies can degrade the results of a meta-analysis. However, at the same time, there is relatively little agreement on the characteristics of methodological

quality [Lipsey and Wilson, 2001, p. 9]. The assessment of methodological quality is, therefore, subject to subjective judgement of researchers.

Additional sources of error can occur as primary studies consulted partly use different definitions and variables [Forza and Di Nuzzo, 1998, p. 839]. As a result, correlations or cross-study effects may be misinterpreted or overlooked when conducting a meta-analysis.

The literature discusses heterogeneity of different studies as another possible source of error of meta-analyses [Lipsey and Wilson, 2001, p. 8]. In particular, the 'apples and oranges' problem is often highlighted by critics. The 'apples and oranges' problem addresses the synthesis of primary studies that deal with different characteristics. As a result, the summarised statistics produced by a meta-analyses might not be powerful [Lipsey and Wilson, 2001, p. 8]. In reality, however, the literature lacks predefined rules to assess an acceptable level of heterogeneity for conducting a meta-analysis. Then again, the assessment of the underlying primary studies is highly subjective. At the same time, Müllner [2002, p. 120] argues that up to a certain level heterogeneity is rather a relevant effect that needs to be described as such, and not as an error.

Publication Bias

As mentioned in Chapter 4 and Section 5.1.2, the publication bias represents an additional potential source of error. The publication bias addresses the fact that statistically significant results with larger mean effect sizes are predominantly published by leading journals and scientists [Lipsey and Wilson, 2001, p. 165]. In contrast, less significant results are often not published in scientific literature. As a result, a high-quality final paper set used for a meta-analysis mostly includes primary studies highlighting significant results and lacks publications with less statistical significance. Then again, researchers must evaluate the trade-off between possible publication bias and adherence to data quality leading to a subjective decision. There is no predefined procedure if the existence of a publication bias is suspected. However, the following two methods are addressed in the literature: Funnel plot and trim-and-fill method.

The funnel plot helps identify the influence of a publication bias on the research results [Egger et al., 1997]. The funnel plot maps the effect size (x-axis) against the precision of the measurement (y-axis) [Jesson et al., 2011, p. 142]. In theory, as a result of smaller sampling errors, larger studies provide better estimates of true effects and vice versa. As shown by Sutton et al. [2000] in Figure 5.5, the funnel plot ideally displays a symmetrical funnel suggesting no publication bias. However, a publication bias cannot be entirely ignored as asymmetries can also be caused by heterogeneity.

Based on the funnel plot, publication bias can be accommodated in a meta-analysis using the trim-and-fill method [Jesson et al., 2011, p. 143]. The trim-and-fill method allows an adjustment of meta-analyses results with the impact of missing studies [Duval and Tweedie, 2000, p. 127]. To accommodate possible publication bias in a meta-analysis, the trim-and-fill method produces a second 'actual' effect size taking into account non-existing studies. However, such a method should only be used if the existence of a publication bias is very likely [Jesson et al., 2011, p. 143]. Figure 5.6 displays a funnel plot including missing studies (filled circles) and the total effect sizes before (open) and after (filled) publication bias adjustment [Hopkins and Smaill, 1999].

The visual examination of publication bias as a result of the funnel plot and the trim-and-fill method can be subjective. Thus, publication bias is additionally examined with an Egger's regression test in the further course [Egger et al., 1997, p. 629].

5 Meta-Analysis

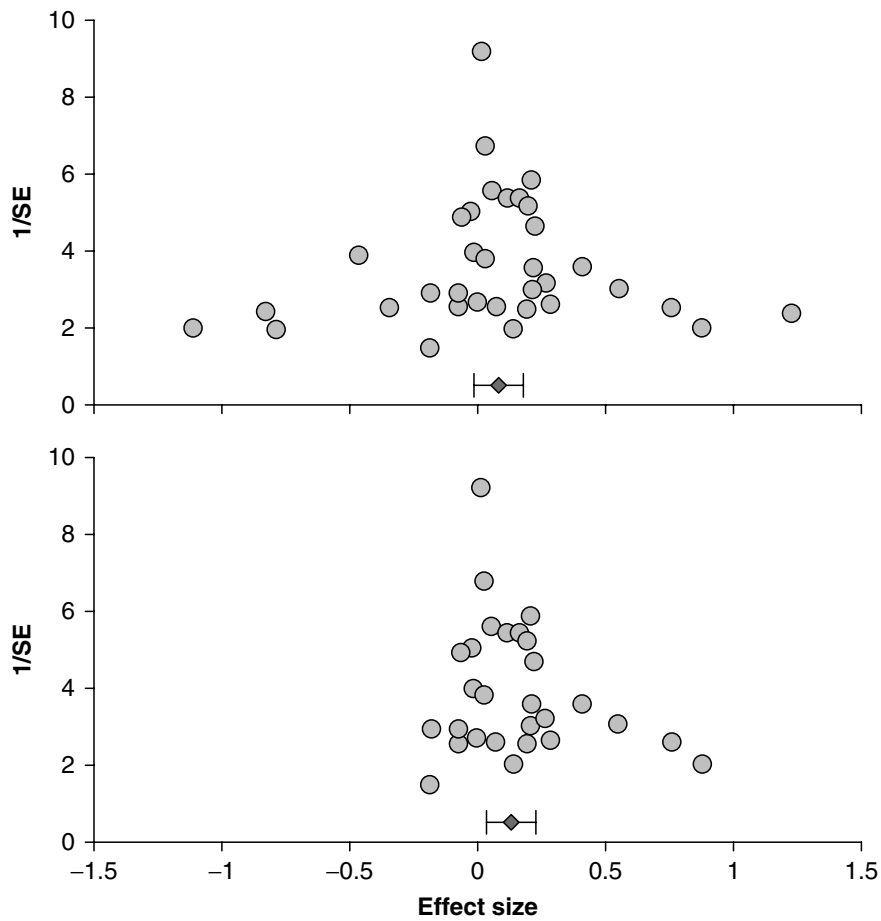


Figure 5.5: Exemplary funnel plot no publication bias (top) and publication bias (bottom) [Sutton et al., 2000, p. 1575]

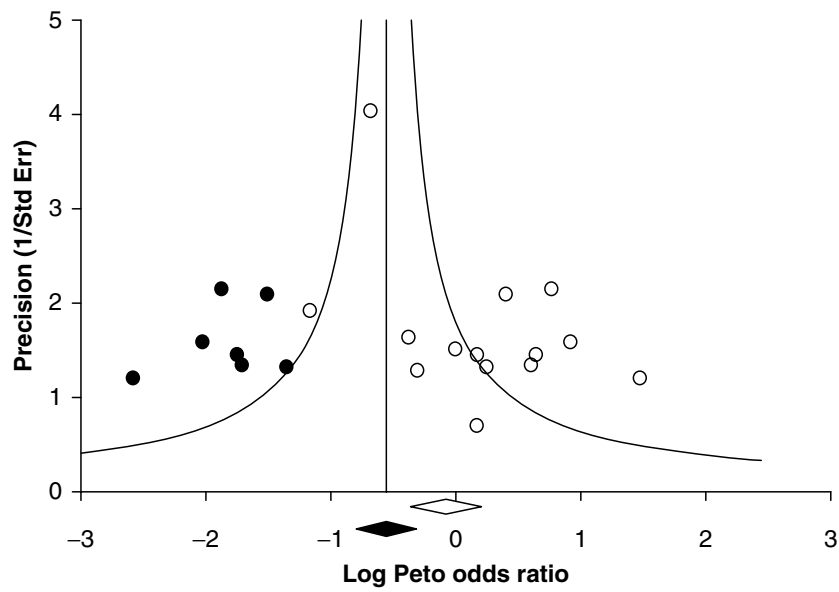


Figure 5.6: Exemplary trim-and-fill method including missing studies [Hopkins and Smail, 1999, p. 1575]

The fact that a classical meta-analysis can only quantify relationships between two variables represents a further limitation [Lipsey and Wilson, 2001]. A classical meta-analysis is not suitable in case that a research question examines multivariate relationships (e.g. in form of a regression analysis).

Lastly, the effort and expertise required to conduct a meta-analysis can be seen as a weakness [Lipsey and Wilson, 2001, p. 7]. In comparison to a traditional qualitative research summary, a meta-analysis requires considerably more time.

5.2 Execution Meta-Analysis

The following sections carry out the meta-analyses according to the procedure described previously. The methods and approaches described above are used to achieve transparent and reproducible meta-analyses.

5.2.1 Formulation of a Research Question

The first step in conducting a meta-analysis is to formulate a research question. As a result, only studies that support the research questions are taken into account in the further course of the meta-analyses. Kate L. Turabian [2013, p. 17f.] distinguishes between three types of questions: Conceptual questions, practical questions, and applied questions.

Conceptual questions help readers to understand a certain problem better and to guide the thoughts [Kate L. Turabian, 2013, p. 17]. Correspondingly, practical questions help develop an approach to change or improve a problematic or improvable situation [Kate L. Turabian, 2013, p. 17f.]. Lastly, applied questions help the readers to first better understand a practical problem before solving it. An applied question helps to develop a step towards the solution of a practical problem [Kate L. Turabian, 2013, p. 18].

This thesis aims to understand the potential of AR solutions in manufacturing environments by conducting three meta-analyses. The underlying problem why AR solutions are considered to support manufacturing activities is described in Section 1.1. This project thus does not address a conceptual question that helps the reader to understand a problem. In reality, the potential and influence of AR solutions must first be researched to develop a concrete procedure to solve the underlying problems. For this reason, the present research question addresses an applied question.

Following Kate L. Turabian [2013], the applied research question of the meta-analyses is as follows:

What influence do AR solutions have on processing time and the error rate, which are a measure of productivity, as well as the cognitive load of workers during manufacturing activities?

The meta-analyses described in the next sections help to answer this question.

5.2.2 Collection of Data

As described in Section 5.1.3, a systematic literature review constitutes the data collection for the meta-analysis. Vom Brocke et al. [2009] suggest the taxonomy by Cooper [1988] for a correct classification of the literature search.

Meta-analyses make use of empirical studies and aim to achieve statistically more powerful assessments. To allow such assessment, the systematic literature review's focus lies on available research outcomes [Cooper, 1988]. This project also aims to 'integrate or synthesize past literature that is believed to relate to the same issue' [Cooper, 1988, p. 108]. At the same time, this project aims to identify central issues in AR applications that have dominated past endeavors. The literature review attempts to represent the influence of AR solutions neutrally. Following Booth [2006], the exclusion criteria do not eliminate a particular point of view. Additionally, conclusions will be based on an exhaustive and selective review [Cooper, 1988, p. 110f.]. The organisation of the systematic literature review follows both a conceptual and methodological approach. Publications that relate to the same abstract ideas and employ similar methods are grouped [Cooper, 1988, p. 111f.]. Lastly, this review intends to address general scholars and practitioners. As a result, the review tries to pay 'greater attention to the implication of the work being covered' [Cooper, 1988, p. 112] than on jargon and details.

Figure 5.7 displays the described taxonomy by Cooper [1988].

Characteristic	Categories			
Focus	Research outcomes	Research methods	Theories	Applications
Goal	Integration		Criticism	Central Issues
Perspective	Neutral Representation		Espousal of Position	
Coverage	Exhaustive	Exhaustive and Selective	Representative	Central
Organization	Historical	Conceptual		Methodological
Audience	Specialized Scholars	General Scholars	Practitioners	General Public

Figure 5.7: Completed taxonomy following Cooper [1988]

Next, a search string is created based on the classification of the literature search by Cooper [1988]. The search string and different combinations of the keywords help to identify relevant publications in the first place. As shown in Figure 5.8, the search string is constructed with three distinct segments: Technology, domain, and the target variable. The corresponding keywords result in 18 individual search strings.

Additionally, the STARLITE methodology is used as a documentation standard (see Figure 5.9) [Booth, 2006]. As a result of an exhaustive and selective sampling strategy, this project considers all literature within predefined boundaries. The search for relevant literature is limited to journal articles and books and is conducted with the help of an evaluation scheme following Vom Brocke et al. [2009]. AR applications have evolved significantly in recent years, and publications have increased considerably since 2014. Consequently, this thesis includes English and German articles between 2014 and 2021. AR solutions are currently implemented in numerous different fields of application. As indicated in Section 1, this research project particularly focuses on the manufacturing industry. For this reason, literature without any empirical evaluation of AR solutions in manufacturing environments is excluded. Seven different databases are chosen not to miss any relevant research outcomes.

3,218 scientific articles meet the predefined criteria as a result of the collection of data phase. Next, the output is evaluated following the evaluation scheme described in Section 5.1.

Technology	Domain	Target variable
<ul style="list-style-type: none"> Augmented Reality Mixed Reality 	<ul style="list-style-type: none"> Manufacturing Maintenance Assembly 	<ul style="list-style-type: none"> Time Error* NASA*

Combination search strings			
#	Technology	Domain	Target variable
1	Augmented Reality	Manufacturing	Time
2	:	:	Error*
3			NASA*
4		Maintenance	Time
5		:	Error*
6			NASA*
7		Assembly	Time
8		:	Error*
9			NASA*
10	Mixed Reality	Manufacturing	Time
11	:	:	Error*
12			NASA*
13		Maintenance	Time
14		:	Error*
15			NASA*
16		Assembly	Time
17		:	Error*
18			NASA*

Figure 5.8: Keywords and search string combinations

S	Sampling Strategy	Exhaustive and selective
T	Type of Studies	Restriction to journal articles, conference papers, and books
A	Approaches	Keyword search in databases, forward search, backward search
R	Range of Years	Consideration of all sources published from 2014 to March 2021
L	Limits	Limitation to English and German sources
I	Inclusion & Exclusion	Focus on empirical user studies of Augmented Reality solutions in manufacturing environments
T	Terms used	Augmented Reality/ Mixed Reality, Manufacturing/ Maintenance/ Assembly, Time/ Error*/ NASA*
E	Electronic Sources	IEEE, ISI Web of Knowledge, Science Direct, Scopus, ACM Digital Library, ProQuest (incl. ABI Informs), JSTOR

Figure 5.9: Completed STARLITE approach following Booth [2006]

5.2.3 Evaluation of Data

As described previously, the identification of relevant studies strongly affects the success of a meta-analysis. Here, it is particularly important to follow a structured approach. This section first presents the results of the literature evaluation in line with the described evaluation scheme (Figure 5.4). Then, the resulting final paper set is classified following step 4 of the framework for systematic literature reviews by Vom Brocke et al. [2009].

The initial keyword search in the mentioned databases results in 3,218 potentially relevant studies. Of these, 249 articles were found on the database IEEE, 93 on ISI Web of Knowledge, 694 on Science Direct, 1,145 on Scopus, 315 on ACM Digital Library, 231 on ProQuest (incl. ABI Informs), and 491 on JSTOR. The various databases partly show considerable overlaps. After removing these duplicates, 1,172 different individual studies remain. With the help of a previously created positive and negative journal list, off-topic journals are sorted out, leaving 994 papers. Similarly, 174 studies remain after a full-title assessment, 76 after an abstract evaluation, and 20 after the final full-text review. Four more relevant studies are identified through a forward/ backward search.

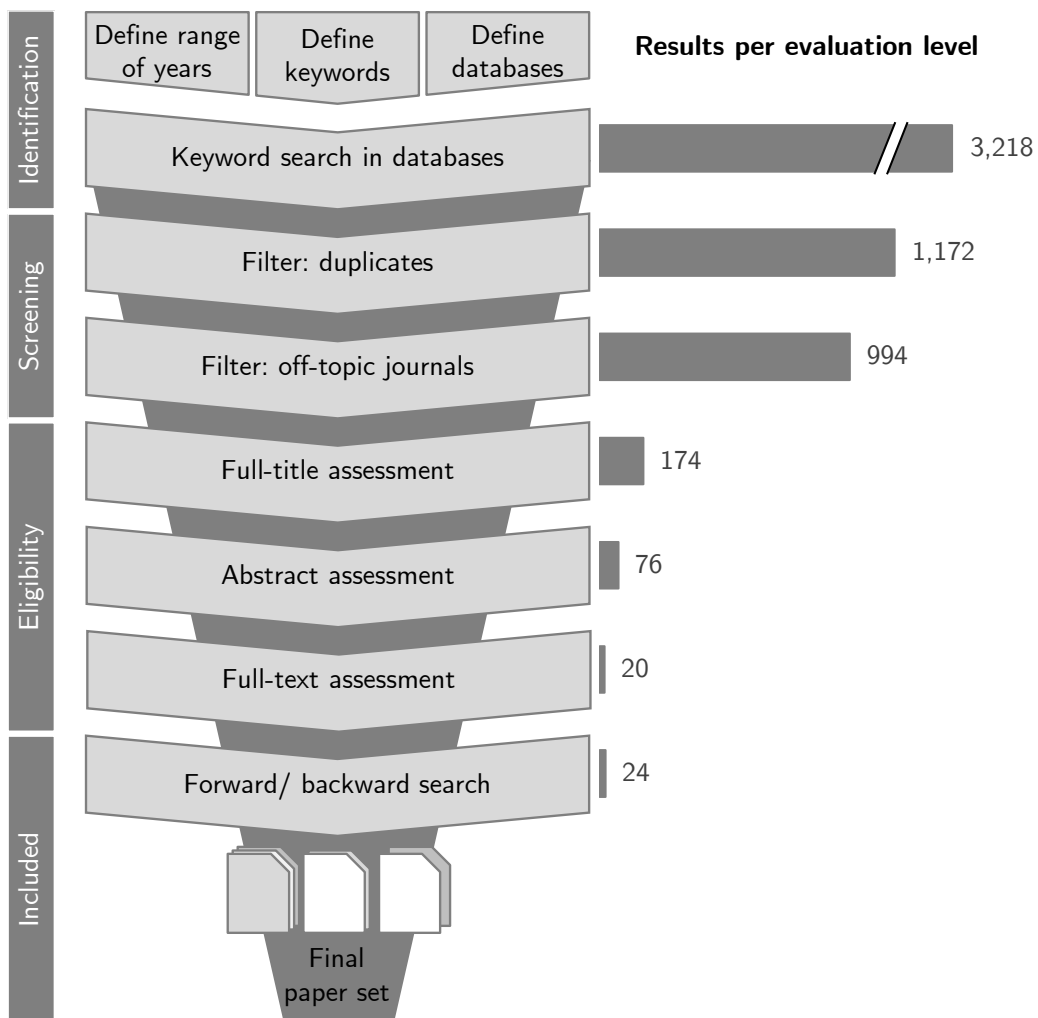


Figure 5.10: Resulting literature per evaluation level

As part of the eligibility assessment, studies are excluded that have not conducted and pub-

lished empirical results of experiments in manufacturing environments. Furthermore, the remaining studies are examined for homogeneity. Here, the domain, the type of data, the research methods, the experimental settings, the number of participants, and the evaluation criteria are considered. Additionally, a comparison between classical instructions and digital assistance technologies is crucial. In the case of single missing data, the authors of the studies were contacted to include the individual studies.

The above-mentioned information of the remaining studies are summarised in Table 5.1 and in more detail in Table B.1 and Table B.2. Particularly the later table constitutes the basis for the following classification and the meta-analyses.

Classification final paper set

In the following, the final paper set is classified based on the aggregated data. In particular, the classification focuses on the distribution of the remaining 24 studies in seven fields of interest.

To begin with, Figure 5.11 displays the distribution of the countries in which the corresponding empirical experiments were carried out. At this point, the relevance of AR in high-wage geographical locations such as Germany, Italy, Austria, Hong Kong, and the United Kingdom is emphasised. The distribution of studies is in line with the fact mentioned above that AR applications can especially strengthen manufacturing processes' competitiveness in high-wage locations [Dachs et al., 2019]. A reason for the geographical focus of scientific research could be that high initial investment costs could pay off more quickly in high-wage countries.

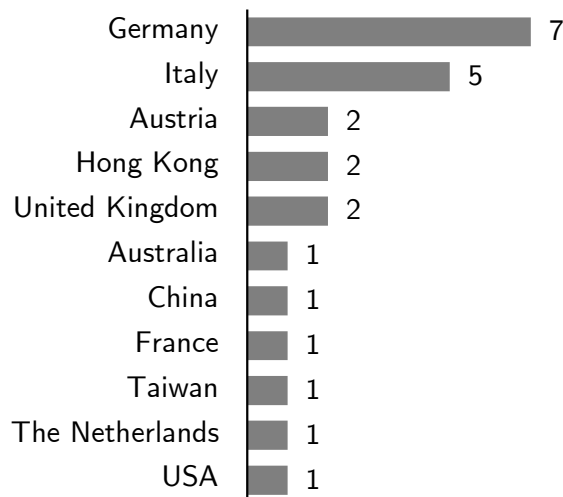


Figure 5.11: Country distribution final paper set

As described in Section 5.2.2, only sources published from 2014 to March 2021 are taken into account for the subsequent meta-analyses. Figure 5.12 illustrates the scientific relevance in recent years. The number of studies conducted increased significantly, especially in 2020. Only a few studies were published before 2014 and do not represent the current technological potential. Correspondingly, the number of available software and hardware solutions in the market has recently also increased at an above-average rate.

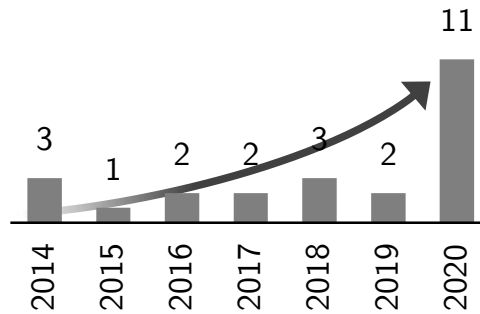


Figure 5.12: Distribution of years final paper set

Next, this section compares the distributions of classical and AR supported instructions. Both in empirical studies but especially in practice, paper instructions are still widespread. Similarly, this is reflected in the distribution of classic instructions shown in Figure 5.13. Besides paper instructions, few studies of the final paper set use PDF tablets or short video instructions with text to provide information classically. Even though paper, PDF tablets, and video instructions are three different mediums, they are comparable. The instructions give very similar information with the help of pictures and text in a non-interactive way. At the same time, the studies as part of the final paper set focus on two enabling technologies of AR, namely HMDs and AR screens. Figure 5.14 illustrates the number of studies using the corresponding technologies. Both technologies are again very comparable as additional information can be displayed in the form of AR. The chosen studies empirically compare the classical and AR supported instructions concerning the described evaluation criteria.

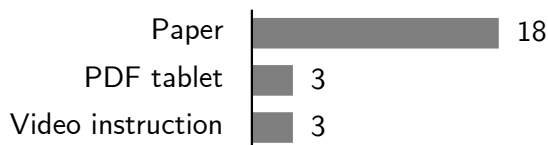


Figure 5.13: Distribution classical instructions

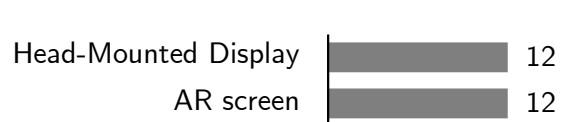


Figure 5.14: Distribution AR supported instructions

Section 2.2.2 describes five different fields of application of AR in manufacturing environments. This meta-analysis particularly focus on the influence of AR solutions on the domains manufacturing, maintenance, and assembly. As shown in Figure 5.15, empirical studies mainly focus on the measurable effects of AR solutions on assembly processes. Kohn and Harborth [2018] have likewise identified a similar focus of research on assembly. At the same time, however, companies are more engaged in projects related to maintenance [Kohn and Harborth, 2018]. Companies only publish quantitative results of their studies in a few cases. All six papers addressing the domain 'maintenance' have been developed in close cooperation with industrial partners.

The experimental settings of the final paper set can be classified into three different categories, namely real product and setting, mock-up, and Lego. The distribution of studies is relatively balanced (see 5.16), real products and settings have found the most applications in the final

paper set. This involves, for example, the assembly of a chainsaw, a car door, and the maintenance of a real machine. The experience of the users performing various tasks differ partly. In contrast, all users had no or very little previous experience working with AR. Individual tasks last between a few seconds and several minutes depending on the experimental setting. Furthermore, the majority of the test persons were either students or employees of a company.

Although the categories of both the domains and the experimental settings appear heterogeneous, the remaining studies have comparable characteristics to enable the meta-analyses. Firstly, the different experimental settings represent simplified activities. These activities include comparable sub-tasks in which items are located, picked, maintained, and assembled. Secondly, the design of the experiments and the usage of the different type of instructions is not familiar for the test persons. Consequently, possible differences between the different categories can be excluded at this stage. However, heterogeneity between the different studies is assessed as part of the statistical analysis in the further course.

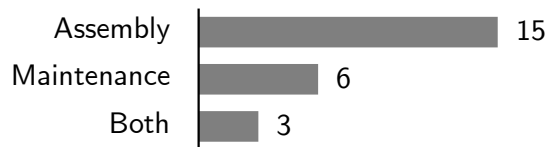


Figure 5.15: Distribution domains of relevant studies

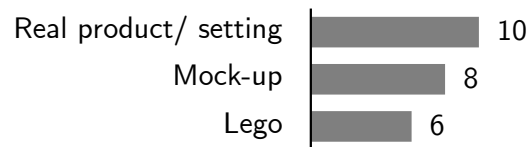


Figure 5.16: Distribution experimental settings




Last, this section illustrates the existence of relevant data for each of the identified evaluation criteria. Almost all studies report the influence of AR applications on the evaluation criteria processing time. Furthermore, the respective percentage changes in the error rates are compared in most studies. The measurement of cognitive load levels requires more effort and knowledge, so this value is only given in 50% of the studies. In addition, time and error rate improvements in industrial applications are more comparable in monetary terms. Lipsey and Wilson [2001, p. 4] mention that the minimum number of comparative studies with the same evaluation criteria for conducting a meta-analysis is three. As shown in Figure 5.17, all three evaluation criteria meet this requirement.






Figure 5.17: Distribution of evaluation criteria

Table 5.1: Overview final paper set

#	Author	Titel	Domain	Experimental setting	Classical instruction	Assistance Technology			
1	Abbas (1)	Impact of Mobile AR System on Cognitive Behavior and Performance during Rebar Inspection Tasks	Maintenance	Real product/setting	Paper	AR screen	X	✓	X
2	Abbas (2)	Impact of Mobile AR System on Cognitive Behavior and Performance during Rebar Inspection Tasks	Maintenance	Real product/setting	Paper	HMD	X	✓	X
3	Blattgerste (1)	Comparing Conventional and AR Reality Instructions for Manual Assembly Tasks	Assembly	Lego	Paper	HMD	✓	✓	✓
4	Blattgerste (2)	In-Situ Instructions Exceed Side-by-Side Instructions in Augmented Reality Assisted Assembly	Assembly	Lego	Paper	HMD	✓	✓	✓
5	Botto	AR for the Manufacturing Industry: The Case of an Assembly Assistant	Assembly	Mock-up	Paper	AR screen	✓	✓	X
6	Brice	AugmenTech: The Usability Evaluation of an AR System for Maintenance in Industry	Maintenance	Mock-up	Paper	HMD	✓	X	✓
7	Büttner	Using Head-Mounted Displays and In-Situ Projection for Assistive Systems – A Comparison	Assembly	Lego	Paper	HMD	✓	✓	X
8	Chu	Comparing Augmented Reality-Assisted Assembly Functions—A Case Study on Dougong Structure	Assembly	Mock-up	Paper	AR screen	✓	✓	✓
9	Fiorentino	Augmented reality on large screen for interactive maintenance instructions	Maintenance	Mock-up	Paper	AR screen	✓	✓	X
10	Gavish	Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks	Maintenance + Assembly	Mock-up	Video instruction	HMD	✓	✓	X
11	Gutsche	Enabling or stressing? – smart information use within industrial service operation	Maintenance + Assembly	Real product/setting	Paper	HMD	✓	X	✓
12	Havard	A use case study comparing AR and electronic document-based maintenance instructions considering tasks complexity and operator competency level	Maintenance	Real product/setting	PDF tablet	AR screen	✓	X	✓
13	Hoover	Measuring the performance impact of using the microsoft HoloLens 1 to provide guided assembly work instructions	Assembly	Mock-up	PDF tablet	HMD	✓	✓	X
14	Hou	Using Animated Augmented Reality to Cognitively Guide Assembly	Assembly	Lego	Paper	AR screen	✓	✓	✓
15	Lampen	Combining Simulation and Augmented Reality Methods for Enhanced Worker Assistance in Manual Assembly	Assembly	Mock-up	Paper	HMD	✓	✓	✓
16	Loch	Comparing Video and Augmented Reality Assistance in Manual Assembly	Assembly	Lego	Video instruction	HMD	✓	X	✓
17	Obermair	Maintenance with AR Remote Support in Comparison to Paper-Based Instructions: Experiment and Analysis	Assembly	Real product/setting	Paper	AR screen	✓	✓	X
18	Pringle	Using an industry-ready AR HMD on a real maintenance task: AR benefits performance on certain task steps more than others	Maintenance	Mock-up	PDF tablet	AR screen	✓	✓	X
19	Sanna	Using Handheld Devices to Support Augmented Reality-based Maintenance and Assembly Tasks	Maintenance + Assembly	Real product/setting	Paper	AR screen	✓	✓	X

#	Author	Titel	Domain	Experimental setting	Classical instruction	Assistance Technology			
20	Uva	Evaluating the effectiveness of spatial AR in smart manufacturing: a solution for manual working stations	Assembly	Real product/setting	Paper	AR screen	✓	✓	✗
21	Wang	Usability evaluation of an instructional application based on Google Glass for mobile phone disassembly tasks	Assembly	Real product/setting	Paper	HMD	✓	✓	✗
22	Werrlich	Comparing HMD-based and Paper-based Training	Assembly	Real product/setting	Paper	HMD	✓	✓	✓
23	Yamaguchi	Video-Annotated Augmented Reality Assembly Tutorials	Assembly	Real product/setting	Video Instruction	AR screen	✓	✓	✓
24	Yang	Comparing the Effects of Paper and Mobile Augmented Reality Instructions to Guide Assembly Tasks	Assembly	Lego	Paper	AR screen	✓	✓	✓

✓ Included ✗ Not included  Time  Error rate  Cognitive load

5.2.4 Analysis and Interpretation of the Data

The studies classified previously form the basis for the analysis and interpretation of the data in this section. For this purpose, the statistically independent effect sizes, summarised in Table B.2, are used. First, the effect sizes are tested for outliers using Grubbs' test [Grubbs, 1969]. Then, this section presents the results from three independent meta-analyses, each focusing on one of the evaluation criteria.

Section 5.1.3 introduces three models for synthesizing the data, namely the equal effect model, the fixed effect model, and the random effect model. The equal effect model does not apply to the remaining data set as the statistical parameters of interest are not the same for every study. Furthermore, clinical and statistical homogeneity cannot be assumed as the studies differ as shown in the classification. Thus, the fixed effect model is not suitable for the analysis of the data. However, the true effect sizes of the selected data differ from each other, so the random effect model is used in this case. Section 5.2.5 performs tests for statistical heterogeneity to check the dispersion of effect sizes.

The meta-analyses are carried out with the software CMA, one of the most widely used software for meta-analyses. The program allows running a wide range of calculations while being easy to use.

Grubbs' test for outliers

This section first assesses the distribution of the effect sizes of the remaining studies. Following Hedges and Olkin [1982, p. 25], outliers can be identified in this way already before the meta-analyses. Above all, a disproportionate influence due to extreme effect size values can be avoided. Grubbs' test helps to identify such outliers for the individual evaluation criteria [Grubbs, 1969].

Following Grubbs' test, the identified relevant studies include two outliers for the evaluation criterion 'errors'. The papers by Büttner et al. [2016] and Sanna et al. [2015] significantly exceed the critical T value provided by Grubbs and Beck [1972]. Nevertheless, both extreme effect size values are not excluded from the meta-analyses. First, the sample sizes are relatively small

compared to the overall number of test persons and do not have a significant impact on the effect sizes [Aguinis et al., 2013, p. 275]. Second, according to Doucouliagos and Stanley [2009, p. 425] and Viechtbauer and Cheung [2010, p. 116], it is unnecessary to exclude potential outliers as the meta-regression is remarkably resilient to high effect sizes. The evaluation criteria 'time' and 'cognitive load' do not include any outliers following Grubbs' test.

Table 5.2 highlights the maximum test statistics for the evaluation criteria in comparison to the critical T values. Additionally, Figure C.1 and Figure C.2 show the individual calculations for the corresponding evaluation criteria in more detail and Grubbs' test statistic is shortly explained in Appendix C.1.

Table 5.2: Overview Grubbs' test for outliers in final paper set

		Time 🕒	Errors ✖	Cognitive load 🧠
1 st run	max G	2.029	2.750	1.789
	critical T *	2.758	2.708	2.412
2 nd run	max G	-	2.773	-
	critical T *	-	2.680	-
3 rd run	max G	-	2.642	-
	critical T *	-	2.652	-

* 5% significance level (two-sided)

Analysis and interpretation

As described previously, all three evaluation criteria are analysed separately using CMA. One of the advantages of CMA is that the software can process studies with different statistical input data. For this reason, in addition to the number of participants and the mean, the standard deviation does not necessarily have to be given, but alternatively the associated p -value.

For each of the three analysis, CMA calculates the effect size, the standard error, the lower and upper limit, the Z-value, and the corresponding p -value. In the following, these statistical results are briefly described.

Effect size. The effect size refers to the magnitude of the difference between an experimental group and a control group. The following analysis make use of Hedges' g to express the effect sizes of the individual studies. Compared to the widely used d by Cohen [2013] and the standard mean difference, Hedges' g reduces biases due to small sample sizes (≤ 20) as is the case for some of the studies [Hedges and Olkin, 1982]. Furthermore, all studies included describe randomized experiments that measure differences between traditional and digitally assisted instructions. Hedges' g is particularly applicable for this case.

Hedges' g indicates the ratio of the difference in means over the pooled and weighted standard deviation:

$$g = \frac{\bar{x}_1 - \bar{x}_2}{s_{pooled}^*}, \quad (5.1)$$

where

$$s_{pooled}^* = \sqrt{\frac{(n_1 - 1) \times s_1^2 + (n_2 - 1) \times s_2^2}{n_1 + n_2 - 2}}. \quad (5.2)$$

According to Cohen [2013, p. 40], magnitudes of g represent the following effects:

- Small effect: $g = 0.20$ to 0.50
- Medium effect: $g = 0.50$ to 0.80
- Large effect: $g = 0.80$ and larger.

If g equals zero, both the classical instructions and the AR assisted instructions are assumed to have equivalent effects. In contrast, a positive value of g indicates beneficial influence of AR, and vice versa.

Standard error. Following Higgins et al. [2019, p. 144], the true effects of measures can never be calculated with absolute certainty. Thus, this meta-analyses make use of the standard errors to express the fuzziness of the values. The standard error depends on two factors in particular, namely the sample size and the variance in the population. The standard error decreases as a result of larger sample sizes or smaller variances.

Lower and upper limit. The confidence interval describes a statistical interval which indicates the precision of the position estimate of a parameter [Sekaran and Bougie, 2016, p. 21]. For a 95% confidence interval, this means that when a study is repeated multiple times in an identical manner, 95% of the results fall between the lower and the upper limit. Heterogeneity can be expected if the confidence intervals of comparable studies do not overlap. In contrast, effect sizes may not be statistically significant if the effect sizes have overlapping confidence intervals. Both indications for heterogeneity and non-significance can be derived visually from a forest plot.

Z-value and p -value. The null hypothesis of a meta-analysis is that the effect size is zero [Jesson et al., 2011, p. 137]. Consequently, no difference between two test groups is expected. The Z-value and the p -value are used to reject this null hypothesis.

The Z-value is an indicator for the significance of the weighted average effect sizes. It is a random variable whose expected value is 0 and whose variance is 1. At the same time, the p -value is an indicator for the probability that the null hypothesis was falsely rejected. Depending on the significance level, the null hypothesis is rejected if the p -value is less or equal to it. Both the effect size and the standard error have a major effect on the calculation of p and Z. In this meta-analysis, the hypothesis is tested for a confidence interval of 95% (two-tailed). Following, the critical Z-values are -1.96 and +1.96.

Figure 5.18 displays the critical values of Z and p and the joint influence on the null hypothesis.

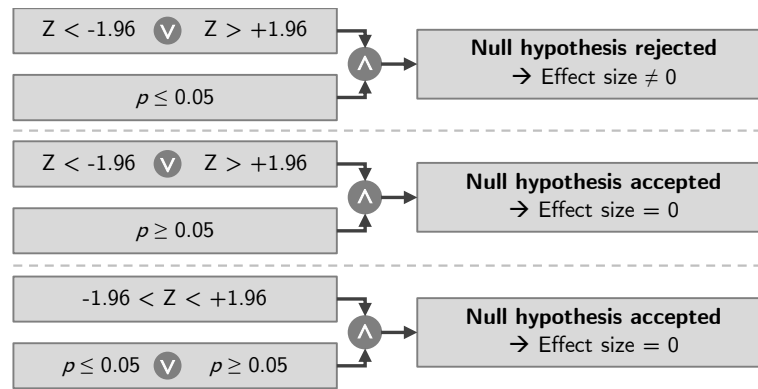


Figure 5.18: Influence Z- and p -values on null hypothesis [Sekaran and Bougie, 2016, p. 382]

Results evaluation criterion 'time'

As described previously, 22 studies are included for the analysis of the evaluation criterion 'time'. The effect sizes of the studies differ in part and reveal both negative and positive effects of AR instructions. Thus, a purely visual evaluation of the effect sizes with the help of the forest plot in Figure 5.19 is not very meaningful.

Nonetheless, the results of the meta-analyses indicate a small effect (Hedges' $g = 0.280$) for the evaluation criterion 'time'. The confidence interval of this evaluation criterion is 0.036 to 0.523. The effect size would fall within this range with a probability of 95% if a study would be repeated. The confidence interval does not include zero, which is why a significant difference between the classical and the AR assisted instructions can be assumed. Inline, Z exceeds the critical limit of +1.96 with a value of +2.251. The corresponding p -value amounts to 0.024. Following Figure 5.18, the null hypothesis can be rejected, and a positive but small effect of AR assisted instructions can be assumed.

Figure C.3 displays a screenshot of the results in CMA. Furthermore, Figure C.4 indicates that the overall effect sizes with one excluded study do not vary strongly, which is in line with the results of Grubbs' test for outliers. The recalculation of the total effect size excluding the study mentioned in the row can be considered a first sensitivity analysis.

Results evaluation criterion 'errors'

For assessing the evaluation criterion 'errors', data from 20 studies in total could be included. Except for two, all studies have a constant effect size greater than zero (see Figure 5.20). Consequently, the effect of using AR assisted instructions can be expected to be positive. The two studies with a negative effect are the previously identified outliers. According to Büttner et al. [2016] and Sanna et al. [2015], the increased error rates in their experiments are a result of technical challenges in dealing with such new technologies.

The initial visual analysis can be strengthened with the statistical results from CMA. Following Cohen [2013, p. 40], the results demonstrate a medium and positive overall effect of AR assisted instructions on the evaluation criterion 'errors' ($g = 0.583$). The confidence interval is 0.302 to 0.864, indicating that the effect size would fall within this range with a 95% probability if a study were repeated. Again, a significant difference between classical and AR supported instructions can be assumed as the interval does not include zero. At the same time, the Z -value

amounts to 4.070, and the corresponding p -value is < 0.001 . Consequently, the null hypothesis, which states that the effect size is zero, can be rejected.

Figure C.5 and Figure C.6 both show a screenshot of the results in CMA. The latter indicates that removing either the study by Büttner et al. [2016] or by Sanna et al. [2015] does not have a significant effect on the overall effect size. The effect size would still be medium even if both studies were removed.

Results evaluation criterion 'cognitive load'

The third meta-analysis aims to assess the influence of AR solutions on the perceived cognitive load of operators during manufacturing activities. Twelve studies are included to determine the overall effect size. As shown in the forest plot in Figure 5.21, the majority of studies reveal a positive influence on cognitive load levels.

Likewise, the calculations in CMA show a small to moderate effect of AR on cognitive load levels of individuals within manufacturing environments. The effect size is 0.325 (standard error of 0.159) and the 95% confidence interval is 0.012 to 0.638. The Z -value of 2.034 is somewhat higher than the critical value of +1.96. In addition, significance is given as a result of p at a level of 0.042. This suggests that AR assisted instructions can potentially decrease cognitive load levels of operators. Users perceived their activities to be significantly less demanding due to the use of AR.

Figure C.7 illustrates the results of the meta-analysis in CMA. The overall effect sizes vary between 0.231 and 0.396 after removing individual studies from the meta-analysis (see Figure C.8). Thus, the data seems to be consistent.

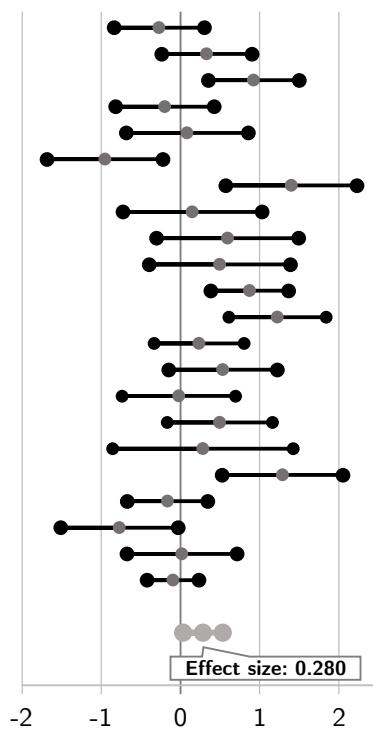


Figure 5.19: Forest plot for 'time'

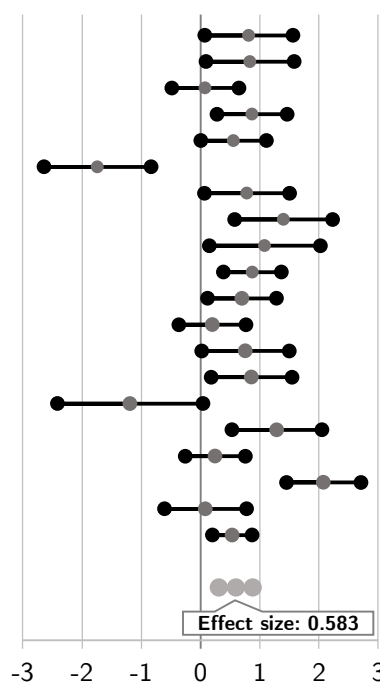


Figure 5.20: Forest plot for 'errors'

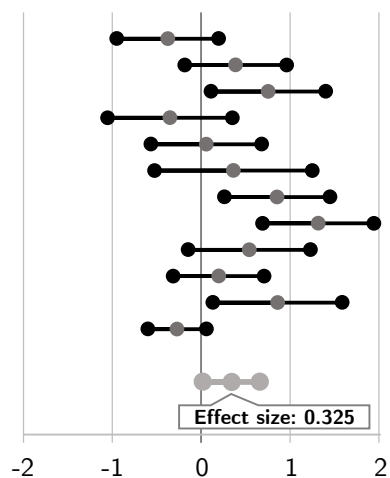





Figure 5.21: Forest plot for 'cognitive load'

Table 5.3 summarises the results of the three meta-analyses. In addition to the effect size, all the values described can be taken from this table. The next section determines the adequacy of the mean effect sizes by analysing heterogeneity and publication bias. Additionally, the impact of moderator variables on the calculated effect sizes is investigated in Chapter 6.

Table 5.3: Statistics meta-analysis

	Time 	Errors 	Cognitive load 
Hedges' g ($\hat{=}$ effect size)	0.280	0.583	0.325
Standard error	0.124	0.143	0.159
Lower limit, upper limit	0.036, 0.523	0.302, 0.864	0.012, 0.638
Z - value	2.251	4.070	2.037
p - value	0.024	0.000	0.042

5.2.5 Performance Sensitivity Analysis

The sensitivity analysis is composed of two independent analysis. First, the individual studies are checked for statistical heterogeneity to interpret the results of the meta-analyses more objectively. Heterogeneity of the effect sizes refers to differences between individual studies [Higgins et al., 2019, p. 259] and is, in this case, assessed using four different indicators. As a result, the comparability of the individual studies included is evaluated.

In the second part, the meta-analyses are examined for publication bias. As described in Section 5.1.5, publication bias refers to the statistically biased presentation of data in science. The underlying reason for this often lies in the preferential publication of positive and significant results. Consequently, the results of a meta-analysis can be distorted.

Heterogeneity

As indicated, heterogeneity describes differences between studies caused by varying measurement methods, populations, and other external influences. Except in perfectly equal experimental settings, heterogeneity always exists. Thus, researchers focus on measuring the impact of heterogeneity on meta-analyses rather than assessing the existence. In general, high levels of heterogeneity between studies can limit the trustworthiness of the overall results [Cochran, 1950].

This thesis carries out a formal assessment of heterogeneity using four different indicators. As a result, the extent to which random variability is responsible for the differences in effect sizes is clarified. In particular, the forest plots provide a graphical indication, and Cochran's Q [Higgins et al., 2019, p. 280], τ^2 by DerSimonian and Laird [1986, p. 180], and I^2 by Higgins et al. [2019, p. 277] allow a quantification.

Forest plots. The forest plots shown in Figure 5.19, 5.20, and 5.21 can provide a first indication for heterogeneity [Fletcher, 2007, p. 95]. Specifically, the focus lies on the positions of the effect sizes and on the width and the overlap of the confidence intervals. Graphical indication of inconsistent effect estimates between individual studies is present if the confidence intervals differ noticeably from each other [Fletcher, 2007, p. 95]. In contrast, homogeneity is present if the 95% confidence intervals overlap for most of the studies.

The three forest plots for the evaluation criteria 'time', 'errors', and 'cognitive load' show varying effect sizes and confidence intervals. Likewise, the confidence intervals of the total effect sizes do not cover the range of the studies included. Thus, heterogeneity can be expected based on the findings from the initial graphical analysis.

Cochran's Q. The literature identifies Cochran's Q as a widely used measure for heterogeneity in meta-analyses [Hardy and Thompson, 1998]. Q refers to the weighted sum of squared differences between each study's results and the overall effect size [Higgins et al., 2003, p. 557]. The weights are equivalent to those used in the meta-analyses. Q is distributed as a chi-square statistic with $k-1$ degrees of freedom [Higgins et al., 2019, p. 280]. In this case, the null hypothesis assumes homogeneity of all studies. A sufficiently small p -value provides insights in the presence of heterogeneity, but not into the present level of it.

It is important to keep in mind that the test results strongly depend on the number of studies included in the meta-analysis. The relevance is reduced as soon as the number of studies is rather small [Gavaghan et al., 2000, p. 421]. For this reason, the relevance of our analyses must also be critically examined.

Cochran's Q is calculated as shown in the following equation:

$$Q = \sum \omega_i (ES_i - \overline{ES})^2 \quad (5.3)$$

The Q-values for all three evaluation criteria exceed the critical chi-square values with $p < 0.001$. Consequently, statistically significant heterogeneity is assumed for each evaluation criterion. As indicated, the extent of heterogeneity cannot be assessed. Additionally, the relevance of Q must be evaluated with caution, given the respective number of studies. Especially for the evaluation criterion 'cognitive load', the number of studies (12) is very low. Table 5.4 provides the exact values of Q.

τ^2 by DerSimonian and Laird. τ^2 by DerSimonian and Laird [1986, p. 180] is often used as another measure for true heterogeneity in meta-analyses. The measure τ^2 calculates the between-studies variance and can directly be calculated from Cochran's Q:

$$\tau^2 = \begin{cases} \frac{Q-df}{C} & \text{if } Q > df \\ 0 & \text{if } Q \leq df \end{cases} \quad (5.4)$$

where C is a scaling factor and is computed as

$$C = \sum w_i - \frac{\sum w_i^2}{\sum w_i} \quad (5.5)$$

Following Huedo-Medina et al. [2006, p. 5], between-studies variance refers to a quantification of the difference between the true population effect sizes in the individual studies. τ^2 can be seen as a first indicator of the level of heterogeneity. However, comparability between meta-analyses that make use of different input formats is not possible [Huedo-Medina et al., 2006, p. 5].

The τ^2 values for all three evaluation criteria indicate true heterogeneity between the studies included. As all three meta-analyses use different input formats, the significance needs to be handled with care.

I^2 by Higgins. I^2 by Higgins et al. [2019, p. 277] is the most widely used measure to quantify the extent of heterogeneity. I^2 expresses the total variance in a meta-analysis and takes into consideration two components, namely random variation and systematic differences between studies [Borenstein et al., 2009]. In comparison to Cochran's Q , I^2 does not depend upon the number of studies considered as can be seen in the calculation using Cochran's Q :

$$I^2 = \frac{Q - df}{Q} \times 100\% \quad (5.6)$$

I^2 is normalised to a range of values between 0 and 100 %. Following [Higgins et al., 2019, p. 259], the different levels of I^2 can be interpreted as follows:

- 0% to 40%: might not be important
- 30% to 60%: may represent moderate heterogeneity
- 50% to 90%: may represent substantial heterogeneity
- 75% to 100%: considerable heterogeneity

In the case of high values of I^2 , conducting a meta-regression is worthwhile Higgins et al. [2019, p. 258]. The aim is to identify moderating variables that explain the heterogeneity, as described in more detail in Chapter 6.

In the case of the described meta-analyses, I^2 ranges between 69.352 and 75.166%. Hence, all three meta-analyses may represent substantial heterogeneity, indicating that almost three-fourths of the observed effects' total variability was caused by true heterogeneity between the studies. Furthermore, it is worthwhile to conduct a meta-regression. The extent of heterogeneity supports the random effects model.

Summing up, the results of the visual analysis, the Q -values, τ^2 , and I^2 indicate that heterogeneity is present in each of the three meta-analyses. As presented previously, the different values of I^2 show that the meta-analyses may represent substantial heterogeneity. The heterogeneity is caused by variance between the studies. Although trust in the evidence of the results is thus reduced, heterogeneity is always present Higgins et al. [2019, p. 259]. Following the heterogeneous effect sizes, it is worthwhile examining various descriptive variables that act as possible moderators (see Chapter 6). One must keep in mind the low number of studies included, especially for the evaluation criterion 'cognitive load'. The results are presented in Table 5.4.

Table 5.4: Heterogeneity statistics for each evaluation criteria ($p < 0.001$ for Q)

	Time 🕒	Errors ✖	Cognitive load 🧠
Point estimate [95% CI]	0.280 [0.036, 0.523]	0.583 [0.302, 0.864]	0.325 [0.012, 0.638]
Q [df]	68.521 [21]	76.509 [19]	38.425 [11]
τ^2	0.224	0.294	0.210
I^2	69.352	75.166	71.373

Publication bias

Next, this section analyses the publication bias described in Section 5.1.5. As suggested by Jesson et al. [2011, p. 142], both a visual assessment of the funnel plot and the trim-and-fill method are examined for each meta-analysis. Asymmetry is also evaluated using the regression test by Egger et al. [1997, p. 629] to eliminate possible subjectivity. The tests are performed using CMA.

Time. First, this section analyses the funnel plot of the evaluation criterion 'time' shown in Figure 5.22. Looking at the distribution of the studies, the funnel plot appears to be symmetric around the effect size. Nevertheless, 8 out of 22 studies are located outside the 95% confidence interval. Following, publication bias cannot be entirely dismissed and needs to be evaluated using additional methods.

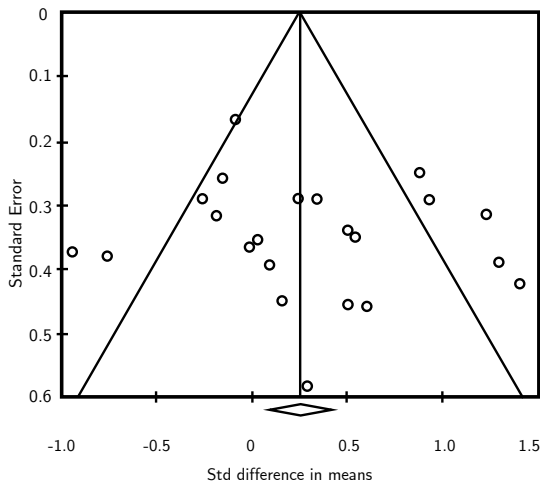


Figure 5.22: Funnel plot for 'time' (random effects model)

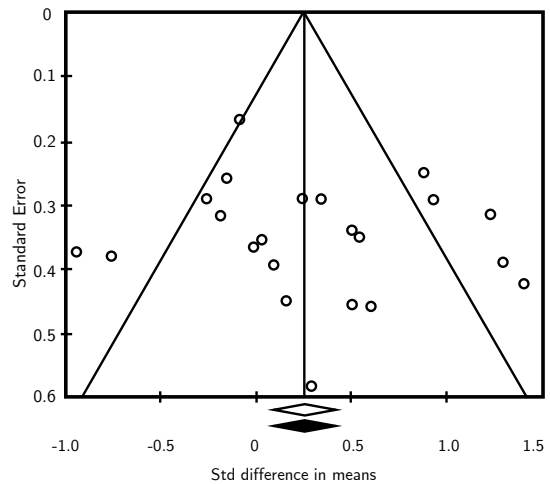


Figure 5.23: Trim-and-fill for 'time' (random effects model)

The trim-and-fill method refutes the results gained from the funnel plot. As can be seen in Figure 5.23, no corrected effect size is calculated in the case of the evaluation criterion 'time'.

Additionally, Table C.1 shows the results of the trim-and-fill method indicating that there is no need for studies to be trimmed. Consequently, publication bias can be neglected in this case [Duval and Tweedie, 2000, p. 127].

Lastly, publication bias is examined with Egger’s regression test. Publication bias is assumed to be present if the intercept is different from zero [Egger et al., 1997, p. 629]. In the case of the evaluation criterion ‘time’, the intercept is 1.22 (see Table C.2). Thus, it is assumed that publication bias is present. However, the p -value amounts to 0.21 (1-tailed) leading to the assumption that the bias is not significant [Egger et al., 1997, p. 629].

Errors. As can be seen in the funnel plot in Figure 5.24, only three studies are located outside the 95% confidence limit for the evaluation criterion ‘errors’. Two of which are the previously identified outliers that are included in the meta-analyses [Duval and Tweedie, 2000, p. 128]. At the same time, the studies in the funnel plot seem to be slightly distorted to the right, which is why publication bias cannot be refused.

This initial evaluation is strengthened by the trim-and-fill method. Figure 5.25 and Table C.1 indicate that there is a need for five studies to be trimmed. Following, publication bias is assumed to be present. Although Figure 5.25 displays an adjusted true effect size, the trim-and-fill method is rather useful to detect publication bias, but not to reliably correct it [Peters et al., 2007, p. 4548].

In contrast, Egger’s regression test reports an intercept of -0.64 and a p -value of 0.36 (1-tailed) (see Table C.2). Again, publication bias is assumed to be present. However, the bias is considered to not be significant for the evaluation criterion ‘errors’.

Summing up, the three methods show that publication bias can be assumed. Nevertheless, Egger’s regression test additionally indicates that it is not significant and can be neglected [Egger et al., 1997, p. 629]. Furthermore, the asymmetry of the funnel plot might not be caused by publication bias, but by the heterogeneity at hand.

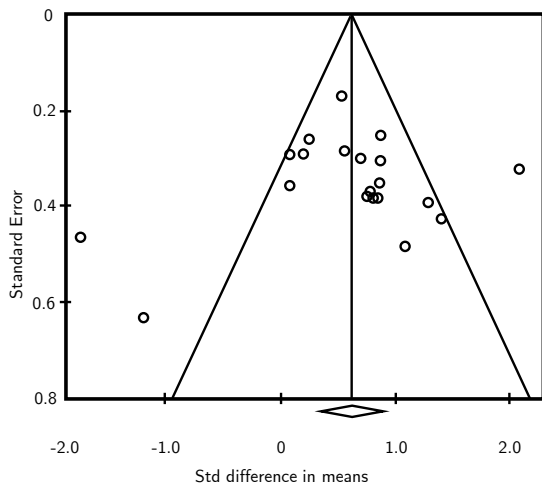


Figure 5.24: Funnel plot for ‘errors’ (random effects model)

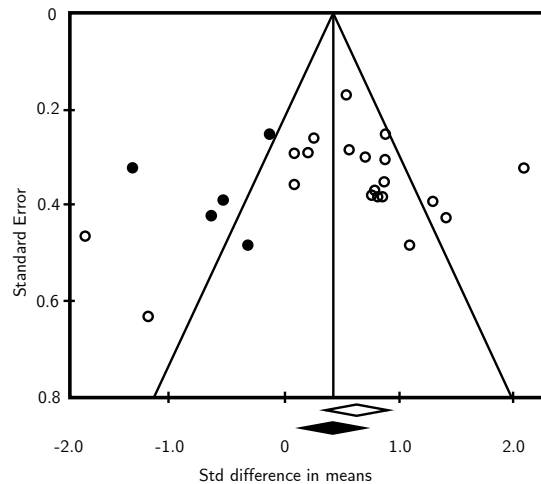


Figure 5.25: Trim-and-fill for ‘errors’ (random effects model)

Cognitive load. Lastly, publication bias is evaluated for the evaluation criterion ‘cognitive load’. The funnel plot in Figure 5.26 illustrates a symmetric distribution of the studies but four

studies are outside the 95% confidence interval. This is a first indication for the presence of publication bias.

The trim-and-fill method in Figure 5.27, however, does not identify any studies with too much effect. Furthermore, no studies need to be trimmed as shown in Table C.1. Consequently, publication bias is considered rather unlikely based on the trim-and-fill method.

Egger's regression test is carried out next. The intercept calculated in CMA amounts to 3.88 and the corresponding one-sided p-value is 0.03. Thus, the intercept deviates significantly from zero and publication bias can be suspected. The power of this method, however, is low with small numbers of studies as in this case [Peters et al., 2006, p. 678]. Detailed results of Egger's regression test can be found in Table C.2.

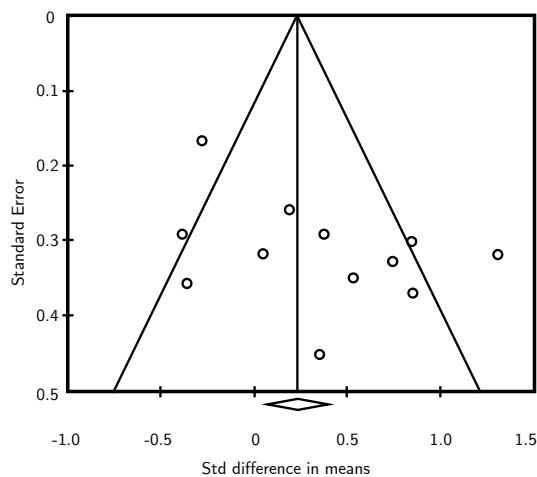


Figure 5.26: Funnel plot for 'cognitive load' (random effects model)

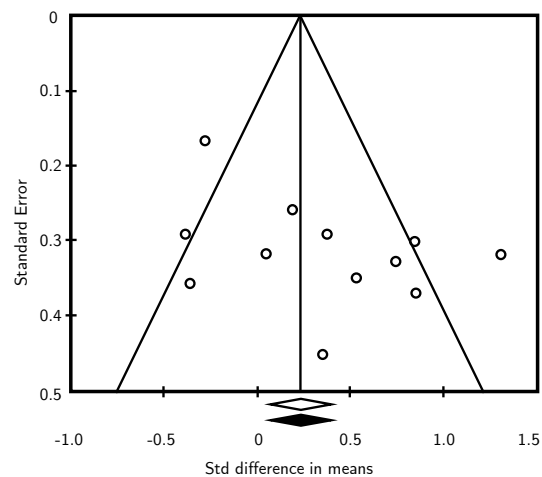


Figure 5.27: Trim-and-fill for 'cognitive load' (random effects model)

In total, different results have emerged in the analysis of publication bias using the funnel plot, the trim-and-fill method, and Egger's regression test for asymmetry. In particular, publication bias is negligible for the evaluation criterion 'time'. In the case of 'errors', the trim-and-fill method indicates publication bias. In contrast, Egger's regression test states that publication bias is negligible. The asymmetry shown by the trim-and-fill method could again be caused by heterogeneity or the outliers included. For 'cognitive load', in contrast to the trim-and-fill method, Egger's regression test indicates publication bias. However, Egger et al. [1997] state that the significance of small study sets must be evaluated in a differentiated way. In general, asymmetry can also be caused by heterogeneity between studies or even randomness [Sterne et al., 2011, p. 2f.].

5.3 Summary Meta-Analysis

Lastly, this section shortly summarises the results of the meta-analyses. As indicated in Section 5.1.3, the described meta-analyses follow six sequential phases in which the influence of AR solutions on three different evaluation criteria is assessed. The collection and evaluation of data are initialised following the taxonomy by Cooper [1988] and the STARLITE approach by Booth [2006]. Based on 18 keyword combinations, a total of 3,218 potentially relevant studies

5 Meta-Analysis

are identified in seven different databases. These are evaluated using an evaluation scheme by Vom Brocke et al. [2009]. In total, the systematic literature review results in 24 studies, of which 22 studies include 'time', 20 include 'errors', and 14 include 'cognitive load' as evaluation criteria.




The meta-analyses conducted based on the literature reviewed indicate that AR solutions indeed have a positive influence on the different evaluation criteria. In particular, AR solutions have a small effect on processing time, a medium impact on error rates, and a small to medium effect on cognitive load during manufacturing activities. In all three meta-analyses, the Z-value is outside the critical limits of -1.96 and +1.96, and the p -value is less than 0.05. Consequently, the null hypothesis, stating that the effect size is zero, is rejected for all three evaluation criteria.

Considering the results of the heterogeneity analysis and the analysis for publication bias, the sensitivity analysis generally confirms the previously analysed results. Nevertheless, the examination for heterogeneity identifies substantial heterogeneity for all three evaluation criteria. As a consequence, trust in the evidence of the results is reduced. Based on these results, the next step is to identify moderator variables (see Chapter 6).

As described earlier, the impact of publication bias on meta-analyses is one of the key weaknesses. In the case of the meta-analyses carried out, Section 5.2.5 shows that there is no evidence for publication bias for the evaluation criterion 'time'. For 'errors' and 'cognitive load' publication bias can not be completely neglected. For both evaluation criteria, publication bias is identified either by the trim-and-fill method or Egger's regression test. The impact of publication bias can hardly be measured.

Table 5.5 presents an overview of the results for the respective evaluation criteria.

Table 5.5: Summary meta-analysis

	Time 	Errors 	Cognitive load 
Effect of AR	Small reducing effect (0.280)	Medium reducing effect (0.583)	Small-medium reducing effect (0.325)
Null hypothesis (effect size = 0)	Rejected	Rejected	Rejected
Heterogeneity (I^2)	May represent substantial heterogeneity (69.521)	May represent substantial heterogeneity (75.166)	May represent substantial heterogeneity (71.373)
Variance between studies (τ^2)	Present (0.224)	Present (0.294)	Present (0.210)
Publication Bias: Trim-and-Fill method	Negligible	Present	Negligible
Publication Bias: Egger's Regressions Test	Negligible	Negligible	Present

6 Meta-Regression

As indicated previously, all three meta-analyses may represent considerable heterogeneity. For this reason, this chapter carries out multiple meta-regressions to be able to describe the existing heterogeneity. First, Section 6.1 provides a definition of meta-regressions. Second, Section 6.2 presents the results of the meta-regressions conducted for each of the evaluation criteria. Lastly, Section 6.3 summarises this chapter.

6.1 Definition Meta-Regression

Meta-regression is a widely used tool to study the relationships between covariates and effect sizes [Huizenga et al., 2011]. This method builds on meta-analytical techniques and linear regression principles to address heterogeneity between studies [Sutton and Higgins, 2008]. The aim is to predict the effect size more accurately according to the values of one or more moderator variables.

As previously presented for the meta-analysis, the meta-regression also distinguishes between three types of models, namely simple regression, fixed effect meta-regression, and random effects meta-regression. The meta-regressions conducted in the next section make use of the random effects regression as a distribution of true effects can be assumed. The model is specified as

$$y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + \dots + \eta + \epsilon_j. \quad (6.1)$$

In this regression equation y_j refers to the effect size in study j , β_0 to the estimated overall effect size, variables x_i ($i = 1, \dots, k$) to different characteristics of the study, η to the variance in studies, and ϵ_j to the between study variation.

One of the advantages of a meta-regression is the possibility of evaluating one or more moderating variables simultaneously. Linear associations between variables and effect sizes and their direction can, hence, be identified easily [Huizenga et al., 2011]. A prerequisite for this, however, is that enough studies are included. Small sample sizes can furthermore lead to bias in a meta-regression. In general, the findings from a meta-regression 'are considered hypothesis-generating' [Baker et al., 2009, p. 1427], but can not be seen as proof of causality.

The meta-regressions for all three evaluation criteria make use of the following five variables available in CMA: p , Q [df], $\tau_{initial}^2$, $\tau_{moderator}^2$, and R^2 . In this case, p presents the statistical significance of the results, Q [df] the chi-square statistics including the degrees of freedom, $\tau_{initial}^2$ and $\tau_{moderator}^2$ the total amount of variance unexplained, and R^2 the amount of variance that can be explained by moderating variables.

In this case, R^2 is of utmost interest as it represents the degree to which a moderator variable can effectively predict a personality trait at a significant level ($p < 0.1$). R^2 refers to the proportion of total unexplained variance that can be explained by the moderator. It is specified as

$$R^2 = \frac{(\tau_{initial}^2 - \tau_{moderator}^2)}{\tau_{initial}^2} \quad (6.2)$$

6.2 Execution Meta-Regression

This section describes the meta-regressions conducted as a result of the presence of heterogeneity. The key parameters described previously are analysed for a total of nine possible moderator variables. A distinction is made between the defined evaluation criteria acting as mutual moderating variables and additional variables related to the characteristics of the underlying studies.

First, this section investigates the evaluation criteria 'time', 'errors', and 'cognitive load' as mutual covariates. Prior to the analysis in CMA, a possible influence of the variables on each other is described below based on different literature.

- **Time:** Following Saptari et al. [2015] and Yang et al. [2010], processing times can have a significant impact on error rates and cognitive load levels of individuals. In particular, Saptari et al. [2015, p. 1201] show that time has a significant effect on errors caused by operators. Especially time pressure increases the error rates significantly, but decreases are possible due to 'unlimited' time. Simultaneously, time may moderate cognitive load levels as they are strongly influenced by it [Yang et al., 2010]. More specifically, time pressure can cause cognitive overload, and underload can occur if processes are very slow [Saptari et al., 2015, p. 1201]. Finally, the experimental setting and supervisors in a real-life environment may also influence 'time' as a covariate. Given these points, the evaluation criterion 'time' may have a moderating effect for 'errors' and 'cognitive load', and the effect is worth investigating.
- **Errors:** Many companies are currently striving for lean production to increase production efficiencies. In this context, it is particularly important to minimise errors since every error involves a time-intensive correction [Abbassinia et al., 2020]. Thus, the number and severity of mistakes an individual makes can impact processing times, and the effect is worth investigating. Simultaneously, errors made by individuals can influence their cognitive load levels. As a result of an error, an operator's confidence can be negatively affected, which in turn can lead to increased mental stress [Wu et al., 2019]. Errors can also be caused by technological errors, which can also influence processing times and cognitive load levels. Given the underlying studies, the meta-regressions can help better understand whether 'errors' drive the measures of effect sizes for both evaluation criteria.
- **Cognitive load:** The evaluation criterion 'cognitive load' could moderate the variance between studies addressing the evaluation criteria 'time' and 'errors'. Lindblom and Thorvald [2014]; Abbas et al. [2020]; Lyell et al. [2018] indicate that cognitive load levels have a direct impact on the performance and error rates of workers in manufacturing environments. In this context, the cognitive load level depends strongly on the availability of information [Lindblom and Thorvald, 2014]. A lack of information can lead to underload, whereas too much information can cause overload, both preventing the operators

from performing at their maximum level (see Figure 6.1). Likewise, Kock [2005, p. 122] indicates that the level of cognitive effort used correlates the evaluation criterion 'time'.

In total, six separate meta-regressions help to identify possible interrelationships between the evaluation criteria. The results of the meta-regressions show that only 'cognitive load' moderates 'time' to a certain extent (see Table 6.1). With $p = 0.0559$, 24% of the total variance is explained by the cognitive load levels. Figure 6.2 displays the linear relationship between the evaluation criteria 'cognitive load' and 'time'. As can be seen, reduced cognitive load levels resulting from the use of AR solutions lead to reduced processing times and vice versa. In this way, the meta-regression reflects the declining course of the grey curve in Figure 6.1 (overload), in which performance is maximised with decreasing cognitive load. The results of Abbas et al. [2020]; Lyell et al. [2018]; Lindblom and Thorvald [2014] can thus partly be transferred to the influence of cognitive load on time as a result of AR applications. Underload is not visible graphically for this meta-regression. None of the other regressions can provide any information on the origin of heterogeneity (see Table 6.1).

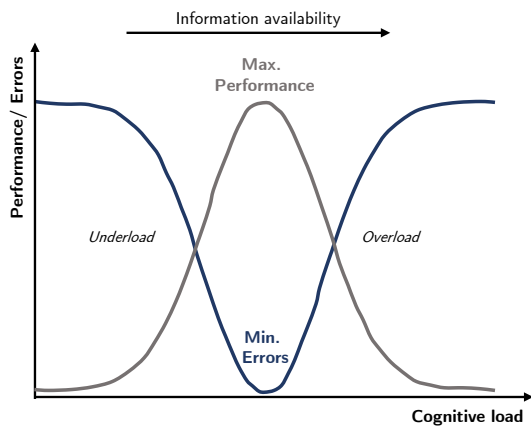


Figure 6.1: Relationship cognitive load, performance, and errors following Abbas et al. [2020]; Lyell et al. [2018]; Lindblom and Thorvald [2014]

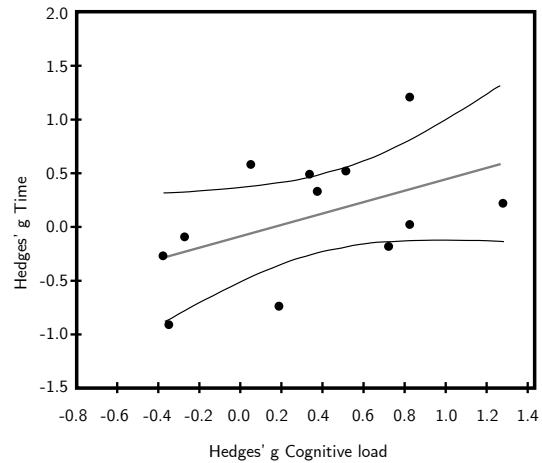


Figure 6.2: Regression 'cognitive load' on 'time'

Table 6.1: Regression statistics evaluation criteria

	Moderator variable	n	Q [df]	p	$\tau^2_{initial}$	$\tau^2_{moderator}$	R ²
Time	Errors	18	0.24 [1]	0.6213	0.265	0.274	0.00
	Cognitive load	12	3.66 [1]	0.0559	0.193	0.147	0.24
Errors	Time	18	0.23 [1]	0.6340	0.327	0.352	0.00
	Cognitive load	8	0.17 [1]	0.6759	0.253	0.320	0.00
Cognitive load	Time	9	0.24 [1]	0.6273	0.149	0.178	0.00
	Errors	8	0.15 [1]	0.7019	0.292	0.357	0.00

6 Meta-Regression

In addition to the moderator variables analysed previously, six further potential variables are identified. The variables are related to the characteristics of the individual studies and include average age, country, year, experimental setting, domain, and assistance technology. Each of the variables are chosen because they could have a plausible impact on the effect sizes and are available for most of the studies. In the following, the potential impact for each of the variables is described in more detail:

- **Average age:** Franke et al. [2019, p. 463] and Wessel et al. [2020, p. 385] show that the affinity for technological interaction decreases with increasing age. The same could apply to the usage of AR applications as an emerging technology. Average age as a moderator variable could thus reveal an influence on the effect sizes for each of the evaluation criteria.
- **Country:** The media company *The Economist* publishes a *Technological Readiness Ranking* at regular intervals. The ranking covers 82 of the largest economies and assesses the economic and political developments that shape technological environments. The better prepared a country is for technological disruptions, the greater the effect sizes for the usage of AR applications could be in that specific environment.
- **Year:** As mentioned earlier, Mizell [2020] has first introduced the term AR in 1992. Since then, technology has continued to develop, and, especially in recent years, attention has increased significantly. Thus, it is worthwhile investigating the influence of the year on the effect sizes.
- **Experimental setting:** As assumed in the meta-analyses, the different experimental settings are comparable with each other. However, heterogeneity between the studies can still be caused by the different experimental settings. For example, participants in Lego experiments could be less distracted in comparison to participants in real life settings.
- **Domain:** As pointed out by Kohn and Harborth [2018, p. 11], the research focus differs between scientists and companies. While scientists tend to focus on the potential of AR in assembly, businesses tend to see more potential in the area of maintenance. The meta-regression is not intended to prove either point of view, but the domain may have an influence on the effect size.
- **Assistance Technology:** Unlike HMDs, tablets have been established in the market for some time. For this reason, it is worthwhile to determine whether the respective assistance technologies moderate the effect sizes of the studies.

The meta-regressions show that only two of the included variables significantly moderate ($p < 0.01$) the effect sizes of two evaluation criteria (see Table 6.2). In the first place, the country moderates the effect sizes of the evaluation criteria 'time'. The results of the meta-regression show that 67% of the variance between studies can be explained by this covariate. In the same way, the publication year moderates the evaluation criterion 'errors'. As displayed in Table 6.2, 78% of the total variance between studies can be explained by the year as a moderating variable. However, in both cases no meaningful explanation on the impact of individual countries and years can be given as highlighted in more detail in Appendix D.

Table 6.2: Regression statistics study characteristics

	Moderator variable	n	Q [df]	p	$\tau^2_{initial}$	$\tau^2_{moderator}$	R^2
Time	Average age	17	1.24 [1]	0.2649	0.205	0.195	0.05
	Country	22	26.88 [9]	0.0015	0.225	0.074	0.67
	Year	22	4.41 [6]	0.6209	0.225	0.252	0.00
	Experimental setting	22	0.38 [2]	0.8288	0.225	0.249	0.00
	Domain	22	0.67 [2]	0.7169	0.225	0.245	0.00
	Assistance Technology	22	1.49 [1]	0.2227	0.225	0.236	0.00
Errors	Average age	14	0.90 [1]	0.3436	0.254	0.239	0.06
	Country	20	0.94 [9]	0.9996	0.295	0.741	0.00
	Year	20	38.76 [6]	0.0000	0.295	0.066	0.78
	Experimental setting	20	2.69 [2]	0.2604	0.295	0.322	0.00
	Domain	20	2.12 [2]	0.3461	0.295	0.302	0.00
	Assistance Technology	20	0.12 [1]	0.7255	0.295	0.322	0.00
Cognitive load	Average age	9	0.58 [1]	0.4444	0.237	0.250	0.00
	Country	12	5.08 [6]	0.5335	0.210	0.206	0.02
	Year	12	8.36 [5]	0.1376	0.210	0.128	0.39
	Experimental setting	12	0.92 [2]	0.6310	0.210	0.215	0.00
	Domain	12	0.50 [2]	0.7782	0.210	0.244	0.00
	Assistance Technology	12	0.18 [1]	0.6721	0.210	0.225	0.00

6.3 Summary Meta-Regression




To summarise, the described meta-regressions determine three covariates that moderate the effect sizes of the studies. These include cognitive load and country as significant moderator variables of the evaluation criterion 'time', and the publication year as a moderator variable of 'errors'. Table 6.3 highlights the results of the meta-regressions and especially focuses on statistically significant results of R^2 . The results of the meta-regression covering the influence of cognitive load on time are in line with studies from other research fields. This meta-regression supports previous research findings, and a statement based on the available data can now also be made for AR applications. Given the statistics of the meta-regressions, heterogeneity of the evaluation criteria can only partly be explained. Especially, an explanation for the evaluation criterion 'cognitive load' is not possible based on the available data.

Additional data, such as the affinity for technological interaction and prior experience, were not reported in either of the chosen studies. This is a field of research and important to investigate

6 Meta-Regression

in future work. Affinity for technological interaction, in particular, is also of great importance in the introduction in the company context to not trigger any further excessive demands on the employees.

Table 6.3: R^2 values of statistically significant moderator variables for $p < 0.1$

	Time 	Errors 	Cognitive load 
Time (Hedges g)	-	-	-
Errors (Hedges g)	-	-	-
Cognitive load (Hedges g)	0.24 ($p < 0.1$)	-	-
Average age	-	-	-
Country	0.67 ($p < 0.01$)	-	-
Year	-	0.78 ($p < 0.001$)	-
Experimental setting	-	-	-
Domain	-	-	-
Assistance Technology	-	-	-

7 Empirical Exploration

The results of the meta-analyses show that AR applications indeed have great potential to reduce processing times, error rates, and cognitive load levels of workers in manufacturing environments. However, industrial companies are partly hesitant to use such technological solutions, and disruption has not yet occurred. To validate the findings from the meta-analyses and the meta-regressions and to obtain further practical knowledge on AR solutions in manufacturing environments, this research project conducts an empirical exploration in the form of expert surveys. Section 7.1 describes the methodological approach of the empirical exploration. Then, a Delphi survey is conducted in Section 7.2. Lastly, Section 7.3 provides a short summary of this chapter.

7.1 Methodological Approach Empirical Exploration

As indicated, this research project aims to validate the findings from the previous analyses and gain additional practical knowledge. For this purpose, a written survey following the Delphi method is conducted in an industrial environment at a consumer goods and chemical company. The Delphi method refers to a multiple-step survey technique developed by the American RAND-Corporation in 1963 and has since established itself as a valuable survey technique [Gordon, 1994, p. 1]. The way in which different Delphi surveys are conducted varies to some extent, but the basic idea is always very similar [Gordon, 1994, p. 1]. The procedure aims to first gather extensive opinions among experts and then evaluate the results with the same or extended group of individuals in further stages. The number of stages and the respective activities in the stages may differ.

The most crucial factor for the success of a Delphi survey is the selection of participants [Gordon, 1994, p. 7]. In particular, the survey results highly depend on the knowledge of the experts and valuable input from them. Participants in a Delphi survey do not necessarily have to be representative of a larger population, and the focus lies on knowledgeable individuals [Galanis, 2018].

Advantages of such a multi-stage survey procedure include the determination of validated forecasts, trends, and opinions by experts in a chosen field of interest [Powell, 2003]. Additionally, more honest and complete answers are expected as the Delphi method is an anonymous survey approach. Creswell and Poth. [2016] emphasise that such surveys can be accessed from different locations in the world and result in fast turnarounds. Furthermore, the recipients have great interest in that specific field of interest and mostly provide thoughtful responses [Keeney et al., 2010].

As shown in Figure 7.1, this paper divides the survey of knowledgeable persons into three phases, namely the preparation phase, 1st round, and 2nd round. First, the initial questionnaire is designed as part of the preparation phase. Based on the initial questionnaire and the corresponding fields of interest, knowledgeable persons can be selected. Next, as part of the 1st

7 Empirical Exploration

round, a group of experts is presented with questions addressing the application of AR. At this stage, the survey consists of open-ended items, and respondents can provide short answers. The results are then consolidated and completed by additional answers based on previous literature research. The collection and consolidation of the answers are then followed by a second round in which participants rank order the extensive collection of responses. Finally, the results of the empirical exploration can be evaluated.

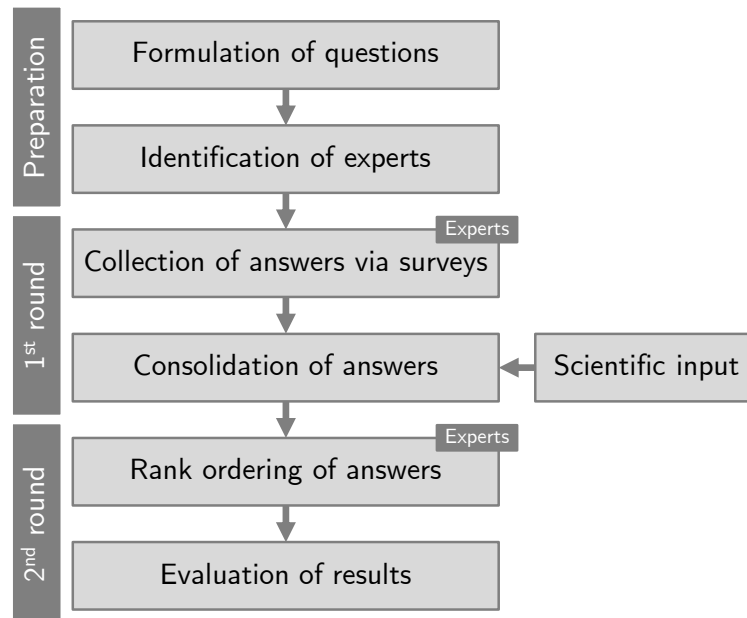


Figure 7.1: Survey approach following the Delphi method

7.2 Execution Empirical Exploration

The following sections highlight the results of the empirical exploration in form of a multi-stage expert survey. Given the survey approach described in the previous section, the first step includes the formulation of survey questions. In this context, this project first defines four areas of interest, namely potentials, challenges, influencing variables, and use cases of AR in manufacturing environments. All of which have great practical but as well scientific relevance and help to fulfill the aims of this empirical exploration. A survey question is derived for each of these areas of interest. The questions are stated in the following:

1. **Potentials of AR:** What potentials can AR applications bring in manufacturing environments?
2. **Challenges of AR:** What challenges do companies face while implementing AR solutions in manufacturing environments?
3. **Influencing variables of AR:** What factors influence the success and acceptance of AR solutions in manufacturing environments?
4. **Use cases of AR:** What relevance do potential use cases of AR in manufacturing environments have?

7.2.1 Participants

The group of experts for this empirical exploration exist of employees working closely related to manufacturing activities at a consumer goods and chemical company. In their research, Mead and Moseley [2001, p. 11f.] identify multiple characteristics of which an expert would need to fulfill at least one. Regarding the expert survey in a corporate context, the experience with the topic and the organisational position are important. Hasson et al. [2000, p. 1011] highlight that in a multi-stage survey procedure, the participants must show commitment and attest their expertise over multiple rounds.

For this study, participants have all been working on or supervising projects related to AR in manufacturing environments and thus qualify as experts. Because of the open-ended nature of the survey questions in the first round, this research project considers eight experts to answer the initial survey. Findings from the literature additionally complement the results. Then, the first round results are rank-ordered by 15 participants qualified as experts in the same way. The experts work in different locations across Europe, including the Netherlands, Spain, Germany, and Italy.

7.2.2 Data Analysis

As described previously, the initial survey in the first round consists of open-ended questions soliciting the potentials, challenges, influencing variables, and use cases of AR in manufacturing environments. The consolidation of the answers was carried out as part of this project and clear consensus among the participants was identified in the first round of the survey. The responses were grouped by topic, and the most meaningful formulations were retained. The expert and literature input yielded 11 consolidated responses for the first question related to the potentials of AR, ten responses related to the challenges, 14 responses related to the influencing variables, and ten responses related to the use cases.

In the second round, 15 participants rank-ordered the responses to determine the most relevant potentials, challenges, influencing variables, and use cases. For the final evaluation, points are attributed according to the individual rankings. The total score is then sorted in ascending order by size (see Tables E.1-E.4).

7.2.3 Results

Although there is less consensus in the second round, clear trends on the relevance of individual responses emerge across the four rankings. For each of the four questions, the three most highly ranked responses are emphasised below. Table 7.1 and Table 7.2 display the entire rankings from the second round. Tables E.1-E.4 show the rankings of individual participants for each of the questions and the corresponding total points given to each response.

Potentials of AR. Following this Delphi survey, the top three potentials of AR in manufacturing environments are (1) faster activities and processes, (2) standardised instructions, and (3) reduction of errors. Faster activities and processes are expected to be achieved through shorter search times and the provision of real-time information whenever needed. Among the participants, 67% rated this answer among their top three potentials, and 44 points are assigned to it. Furthermore, in an industrial environment, standardised instructions are expected to be highly beneficial as consistent quality outputs and fewer dependencies on experienced workers can be

7 Empirical Exploration

a crucial competitive advantage (45 points). Third, the reduction of errors can result from more targeted provision of real-time information and by immediate and automatic error analysis (55 points).

The described findings align with the empirical results from previous meta-analyses on the evaluation criterion "time" and "errors". At the same time, however, AR experts working in this specific consumer goods and chemical company do not expect the reduction of cognitive load levels to be of high importance (see Table 7.2 and Table E.1).

Challenges of AR. Companies still face various challenges during the implementation of AR solutions in manufacturing environments, and large-scale diffusion has not yet taken place. The top three challenges in the consumer goods and chemical company are (1) lack of proven business models with positive return on investments, (2) data protection, privacy, and security issues, and (3) technological readiness. During the second round of this Delphi survey, 73% of the AR experts ranked missing business models among the top three challenges. The complexity of building and using an AR experience for individual use cases is only in a few cases in proportion with the potentials gained. Additionally, data protection, privacy, and security requirements are especially high in large and globally active companies. According to 53% of the participants, this area of interest still shows significant room for improvement. Lastly, 47% of the AR experts assess technological readiness to be one of the biggest three challenges. In particular, the battery life of AR enablers, the viewing angles, and the robustness are still limited.

The findings from this empirical exploration are in line with the results of further research projects. Among others, Kohn and Harborth [2018, p. 12] and Danielsson et al. [2020, p. 1301] as well highlight data protection, privacy, and security issues, and lack of technological readiness as the key challenges of AR in manufacturing environments.

Influencing variables of AR. The success and acceptance of AR solutions in manufacturing environments highly depend on various variables. The following top three out of 14 influencing variables can be derived from this empirical exploration: (1) Technical affinity of operators and management, (2) ease of use of AR applications, and (3) reliability of the technology. Appearing in 67% of the top three rankings and rated with 44 points, technical affinity of users is of utmost importance to successfully implement such emerging technology. Almost as highly prioritised (60% top three rankings, 61 points), the easiness of AR applications significantly influences the success and acceptance. Third, the reliability of such technology is expected to be present already at an early stage of the implementation to keep users' motivation high.

An overview of all other ranked influencing variables can be found in Table 7.2. The ranking emphasises that multiple additional moderator variables could be examined in future meta-regressions.

Use cases of AR. As described previously, the biggest challenge during the implementation of AR in the manufacturing environment of this particular consumer goods and chemical company lies in the identification of profitable use cases. Here, in the short term, 87% of the experts surveyed expect great relevance of AR in training (31 points). Among other training applications, safety, change-over, maintenance, and assembly training are of utmost relevance. Furthermore, the relevance of training applications in an industrial setting is closely followed by virtual collaboration and remote maintenance (32 points). Among the AR experts at the consumer goods and chemical company, 80% rated this use case among their top three relevant use cases. Next, 52 points are assigned to task-guidance applications on-the-job. The technological readiness of AR enablers prevents task guidance use cases to be of higher relevance at

the current state. However, future technological developments can support on-the-job applications. Task guidance applications can as well include safety, change-over, maintenance, and assembly instructions.

Table 7.1: Results empirical exploration *potentials* and *challenges*





Potentials 	Challenges 
<ol style="list-style-type: none"> 1. Faster activities/ processes 2. Standardised instructions (same quality output) 3. Reduction of errors/ increased reliability 4. Independent and flexible instructions 5. Increased safety 6. Better internal and external communication 7. Better problem identification and description 8. Hands-free provision of information 9. Less cognitive load of users 10. Support of home-office 11. Social employer branding 	<ol style="list-style-type: none"> 1. Lack of proven business models with positive ROI 2. Data protection, privacy, and security 3. Technological readiness 4. Acceptance from users 5. Limited workplace safety and health 6. Missing identification of processes/ use cases that are ready to implement AR 7. High initial and running costs 8. Compatibility with different technologies 9. Lack of development resources 10. Lack of AR app design and development standards

Table 7.2: Results empirical exploration *influencing variables* and *use cases*

Influencing variables 	Use cases 
<ol style="list-style-type: none"> 1. Technical affinity of operators and management 2. Ease of use of AR applications 3. Reliability of technology 4. Change management 5. Leadership commitment 6. Use case readiness 7. Degree of digitisation in the manufacturing environment 8. Comfort and aesthetics of devices 9. Corporate culture incl. curiosity of employees 10. Age 11. Compatibility with different technologies (AI, IoT, etc.) 12. Documentation of instructions and workshops 13. In-house app development skills 14. Cultural background 	<ol style="list-style-type: none"> 1. Trainings (pre-job)¹ 2. Virtual collaboration/ Remote maintenance 3. Task guidance (on-the-job)¹ 4. Task validation (post-job)¹ 5. Virtual factory planning 6. Audits 7. Plant visits 8. Presentation of IoT data 9. Navigation in plants 10. Material flow visualization <p><small>¹: individual use cases related to safety, change-over, maintenance tasks, assembly tasks, material handling, machine parameter adjustment, etc.</small></p>

7.3 Summary Empirical Exploration

Lastly, this section shortly summarises the results of the empirical exploration. As indicated in Section 7.1, a Delphi survey was conducted as part of this research project. The survey addresses the potentials, challenges, influencing variables, and use cases of AR in manufacturing environments. In the context of this empirical exploration, a total of 15 AR experts working in a consumer goods and chemical company were interviewed.

The results from this empirical exploration partially support the findings of the meta-analyses and the meta-regressions. Following the AR experts, there is a great potential of AR in manufacturing environments, and multiple use-cases can be implemented. However, both industrial companies and universities still face numerous challenges to enable large-scale diffusion of AR in manufacturing environments, and the acceptance is highly dependent on various influencing variables that need to be kept in mind. Section 7.2 provides an overall ranking for each of the four fields of interest, and the top three results are outlined correspondingly.

8 Discussion and Conclusion

Finally, this chapter includes the last two sections of this thesis. First, Section 8.1 discusses the results of this thesis and the applicability to the research objective. Then, the discussion is followed by a conclusion of the thesis including a brief summary of the findings.

8.1 Discussion

The main objective of this research was to understand if and how the use of AR solutions can benefit manufacturing activities. The results of the meta-analyses show that AR application indeed have a small to medium positive effect on the evaluation criteria 'time', 'errors', and 'cognitive load'. In addition, this thesis explores potential interrelationships between these factors and puts the results into the context of a chemical and consumer goods company. In the following, this section discusses the main findings covering the scientific and practical relevance, possible limitations of this research project, and future research suggestions.

Scientific relevance

Individual studies in the scientific literature show ambiguous results on the effects of AR applications in manufacturing environments. In particular, a statistically powerful empirical assessment of the impact was still missing. The present study, however, hypothesized that AR solutions have a positive effect on (1) processing times, (2) error rates, and (3) cognitive load levels of workers during manufacturing activities. By synthesising several small studies into one large study, the present meta-analyses provide more powerful statistical proof of the hypotheses and thus close the knowledge gap. Furthermore, this thesis identifies potential moderating variables and examines the interrelationships using multiple meta-regressions. This section discusses the scientific relevance and the implications of the obtained results for each of the three hypotheses.

This research identifies 'time' as the most relevant evaluation criterion in existing literature. Overall, the results of the meta-analysis shows a small reducing effect of AR applications on processing times. As a result, the described scientific knowledge gap for the evaluation criterion 'time' can be closed. Based on the underlying studies the positive effects of higher media naturalness described by Kock [2005] can now also be transferred to AR applications. Furthermore, a statistical correlation between cognitive load and time was identified for the use of AR solutions. These findings are backed up by neuroscientific research published by Abbas et al. [2020], Lyell et al. [2018], and Lindblom and Thorvald [2014].

Besides 'time', the 'error rate' is a crucial parameter in production environments and scientific literature. Results of the meta-analysis shows a significant medium reducing effect on the evaluation criterion 'error rate'. In addition, the positive impact on error rates is also strengthened by expert interviews. Hence, the results obtained in the meta-analysis both statistically confirm

the second hypothesis and expand current scientific studies addressing the positive effects of AR. No significant correlations have been found between 'errors' and the evaluation criteria 'time' and 'cognitive load'. Instead, the meta-regressions identify the year of publication as a moderator variable of the effect sizes, bringing a new potential for research as described in the further course of this section.

Finally, the third hypothesis addresses the positive effect of AR solutions on cognitive load. Although fewer studies measure and report cognitive load, the results reveal a significant small to medium reducing effect of such technology. The previously performed studies individually provide low significance, which the synthesis can clearly improve into one large study. The results obtained from the meta-analysis are supported by experts who indicate that cognitive skills and affinity for technological interaction are particularly important for the success of AR applications. For 'cognitive load', no significant covariates have been found. However, as described previously, the level of cognitive load impacts the evaluation criterion 'time'. Given these results, hypothesis three on the effect of AR applications on cognitive load levels can be accepted.

Practical relevance

Manufacturing companies are confronted with challenges due to increasing flexibility requirements and skill gaps. This research shows that AR applications offer an efficient way to overcome these tensions by enhancing the interaction between people and technology. In particular, the meta-analyses provide a statistically powerful empirical assessment from which the insights can be used in industrial environments. Based on the findings of this research, companies can decrease processing times, error rates, and cognitive load levels by using AR. In line with the media naturalness theory by Kock [2005], AR applications are a great way to enhance communication and knowledge sharing compared to widely used static instructions. These findings are strengthened and extended by the results of an empirical exploration in a consumer goods and chemical company.

Challenges during the introduction of AR applications are often not discussed in the literature. However, large-scale diffusion of AR solutions in industrial environments has not yet taken place. The present research identifies multiple challenges that prevent companies from successfully implementing such technology. These insights can be used to counteract potential challenges at an early stage by developing countermeasures and taking special care of them.

As highlighted in previous research and the empirical exploration, companies are still struggling to find suitable and especially profitable use cases of AR to leverage the potentials. This research shows different use cases according to their potentials following the knowledge of experts in that field. Nevertheless, companies need to match potential use cases to their weaknesses to improve their competitiveness. For instance, a company with relatively long search times due to ignorance should build an AR application that supports the operators during navigation in the factory.

Limitations

The aforementioned results present multiple important limitations that need to be kept in mind. To begin with, the meta-analysis itself reveals methodological limitations. First, the results of the meta-analysis highly depend on the quality of the underlying studies. Furthermore,

a subjective judgement of researchers to include individual studies can affect the calculated effect sizes. However, the present work has addressed this limitation in the best possible way through a structured evaluation scheme.

The meta-analyses show that AR solutions have a positive effect on the evaluation criteria 'time', 'errors', and 'cognitive load'. Especially for the evaluation criterion 'cognitive load' the number of studies included is comparably small. As a consequence, the significance of the statistical tests that examine possible biases can be negatively influenced. Even if publication bias and heterogeneity were partially demonstrated, it is not possible to determine what impact these have on the results of the meta-analyses.

Although heterogeneity is suspected, the variance cannot be fully explained using multiple meta-regressions. Meta-regressions are limited to the information available. In particular, the present studies do not report potential moderator variables such as prior experience with AR and the affinity for technological interaction. An expanded data set would allow differences between studies to be further explored.

Lastly, only a small group of experts from a specific industry is interviewed as part of this research. Consequently, the knowledge gained can not yet be applied across industries and manufacturing companies. However, the results can serve as an initial guide.

Future research

As shown previously, AR solutions are gaining more and more interest in both scientific research and in industries. Therefore, recommendations for future research are described in the following. First of all, a repetition of the meta-analyses would be necessary as soon as a sufficient number of new empirical studies on the topic of AR solutions have been published. Further empirical studies could include Electroencephalography (EEG) testing in addition to the NASA-TLX test. This would allow the results to be validated with a larger number of studies based on other scientific methods.

Furthermore, a long-term validation in an industrial environment is still missing. When planning and conducting such validation, it must be considered that in countries with high data protection requirements, tests in the company environment are difficult to achieve.

The identified variance between the studies can not be fully explained, and additional moderator variables might exist. Thus, it is worth including further possible covariates and running additional meta-regressions. The insights gained could be of great scientific and practical relevance.

Finally, there is a need for further research on the potentials, challenges, influencing variables, and use cases of AR solutions in manufacturing environments across industries. Additional surveys and interviews with experts will need to be conducted to gain these insights.

8.2 Conclusion

Manufacturing companies are undergoing major changes in today's world of globalization and digitization. Among other challenges, companies - particularly in high-wage countries - are facing growing competition and disruptive market changes. To counter these challenges in manufacturing environments, AR solutions are a promising technology. The research objective

8 Discussion and Conclusion

of this study was to determine if and how the use of AR solutions can benefit manufacturing activities. In order to answer this research question, three meta-analyses are conducted addressing the most prominent evaluation criteria 'time', 'errors', and 'cognitive load'.

The meta-analyses reveal that based on the present studies AR applications indeed have a positive effect on all three evaluation criteria. Hence, the defined hypotheses are all accepted as displayed in Table 8.1.

Table 8.1: Results of hypotheses

#	Hypothesis	Result
1	AR solutions have a reducing effect on processing times of workers during manufacturing activities.	Accepted
2	AR solutions have a reducing effect on error rates of workers during manufacturing activities.	Accepted
3	AR solutions have a reducing effect on cognitive load levels of workers during manufacturing activities.	Accepted

What is more, this research project identifies 'cognitive load' as a moderator variable for the evaluation criterion 'time'. The variance between studies can be partly explained by different cognitive demands. These findings are in line with prior research conducted in the field of neuroscience [Kock, 2005].

Finally, the results of the meta-analyses and of the meta-regressions are supported by industrial experts working in a consumer goods and chemical company. With the help of ranked answers in the four categories potentials, challenges, influencing variables, and use cases, the relevance of AR is also be emphasised in an industrial environment.

Bibliography

- Abbas, A., Seo, J., and Kim, M. (2020). Impact of Mobile Augmented Reality System on Cognitive Behavior and Performance during Rebar Inspection Tasks. *Journal of Computing in Civil Engineering*, 34(6):04020050.
- Abbassinia, M., Kalatpour, O., Motamedzadeh, M., Soltanian, A., and Mohammadfam, I. (2020). The Application of Lean Production in Reducing Human Error and Improving Response in Emergencies: A case Study in a Petrochemical Industry. *Journal of Ergonomics*, 8(2):39–49.
- Aguinis, H., Gottfredson, R. K., and Joo, H. (2013). Best-Practice Recommendations for Defining, Identifying, and Handling Outliers.
- Allen, M. (2017). *The SAGE Encyclopedia of Communication Research Methods*. SAGE Publications.
- Azuma, R. T. (1997). A Survey of Augmented Reality. *Presence: Teleoperators and Virtual Environments*, 6:355–385.
- Baker, W. L., Michael White, C., Cappelleri, J. C., Kluger, J., and Coleman, C. I. (2009). Understanding heterogeneity in meta-analysis: The role of Meta-regression. *International Journal of Clinical Practice*, 63(10):1426–1434.
- Baroroh, D. K., Chu, C. H., and Wang, L. (2020). Systematic literature review on augmented reality in smart manufacturing: Collaboration between human and computational intelligence. *Journal of Manufacturing Systems*.
- Basavanna, M. (2015). Research Methods in Psychology. *Psychology for Nurses*, pages 27–27.
- Bijmolt, T. H. and Pieters, R. G. (2001). Meta-Analysis in Marketing when Studies Contain Multiple Measurements. *Marketing Letters*, 12(2):157–169.
- Booth, A. (2006). "Brimful of STARLITE": Toward standards for reporting literature searches. *Journal of the Medical Library Association*.
- Borenstein, M., Hedges, L., Higgins, J., and Rothstein, H. (2009). *Introduction to meta-analysis*. John Wiley & Sons, Ltd.
- Burggräf, P., Dannapfel, M., Adlon, T., Riegauf, A., and Schmied, J. (2020). Automation configuration evaluation in adaptive assembly systems based on worker satisfaction and costs. In *Advances in Intelligent Systems and Computing*, volume 959, pages 12–23. Springer Verlag.
- Büttner, S., Funk, M., Sand, O., and Röcker, C. (2016). Using head-mounted displays and in-situ projection for assistive systems - A comparison. In *ACM International Conference Proceeding Series*, volume 29-June-20. Association for Computing Machinery.
- Chen, B., Wan, J., Shu, L., Li, P., Mukherjee, M., and Yin, B. (2017). Smart Factory of Industry 4.0: Key Technologies, Application Case, and Challenges. *IEEE Access*, 6:6505–6519.

BIBLIOGRAPHY

- Chen, C. H., Lee, I. J., and Lin, L. Y. (2015). Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders. *Research in Developmental Disabilities*, 36:396–403.
- Cochran, W. G. (1950). The comparison of percentages in matched samples. *Biometrika*, 37(3-4):256–266.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.
- Cooper, H. M. (1988). Organizing knowledge syntheses: A taxonomy of literature reviews. *Knowledge in Society*, 1(1):104–126.
- Creswell, J. W. and Poth., C. N. (2016). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches*. SAGE publications, Inc.
- Dachs, B., Kinkel, S., and Jäger, A. (2019). Bringing it all back home? Backshoring of manufacturing activities and the adoption of Industry 4.0 technologies. *Journal of World Business*, 54(6).
- Daft, R. L. and Lengel, R. H. (1986). Organizational Information Requirements, Media Richness and Structural Design. *Management Science*, 32(5):554–571.
- Danielsson, O., Holm, M., and Syberfeldt, A. (2020). Augmented reality smart glasses for operators in production: Survey of relevant categories for supporting operators. In *Procedia CIRP*, volume 93, pages 1298–1303.
- David Greenfield (2017). Augmented Reality for Industry. <https://www.automationworld.com/products/software/article/13315547/augmented-reality-for-industry>.
- DerSimonian, R. and Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, 7(3):177–188.
- Doucouliafos, H. and Stanley, T. D. (2009). Publication selection bias in minimum-wage research? A meta-regression analysis. *British Journal of Industrial Relations*, 47(2):406–428.
- Duval, S. and Tweedie, R. (2000). Trim and Fill: A Simple Funnel-Plot-Based Method. *Biometrics*.
- Egger, J. and Masood, T. (2020). Augmented reality in support of intelligent manufacturing A systematic literature review.
- Egger, M., Smith, G. D., Schneider, M., and Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315(7109):629–634.
- ElMaraghy, H., Schuh, G., Elmaraghy, W., Piller, F., Schönsleben, P., Tseng, M., and Bernard, A. (2013). Product variety management. *CIRP Annals - Manufacturing Technology*, 62(2):629–652.
- Feldmann, K., Geiger, M., and Lang, S. (2007). *Durchgängige Mitarbeiterinformation zur Steigerung von Effizienz und Prozesssicherheit in der Produktion*. Meisenbach.
- Fletcher, J. (2007). What is heterogeneity and is it important? *British Medical Journal*, 334(7584):94–96.

- Forza, C. and Di Nuzzo, F. (1998). Meta-analysis applied to operations management: Summarizing the results of empirical research. *International Journal of Production Research*, 36(3):837–861.
- Franke, T., Attig, C., and Wessel, D. (2019). A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human-Computer Interaction*, 35(6):456–467.
- Gaimon, C. and Singhal, V. (1992). Flexibility and the choice of manufacturing facilities under short product life cycles. *European Journal of Operational Research*, 60(2):211–223.
- Galanis, P. (2018). The Delphi method. *Archives of Hellenic Medicine*, 35(4):564–570.
- Galaske, N. and Anderl, R. (2016). Approach for the development of an adaptive worker assistance system based on an individualized profile data model. In *Advances in Intelligent Systems and Computing*, volume 490, pages 543–556. Springer Verlag.
- Gavaghan, D. J., Moore, R. A., and McQuay, H. J. (2000). An evaluation of homogeneity tests in meta-analyses in pain using simulations of individual patient data. *Pain*, 85(3):415–424.
- Gong, G., Deng, Q., Gong, X., Liu, W., and Ren, Q. (2018). A new double flexible job-shop scheduling problem integrating processing time, green production, and human factor indicators. *Journal of Cleaner Production*, 174:560–576.
- Gordon, T. J. (1994). The Delphi method. *Futures research methodology*, 2.3:1–30.
- Gröger, C., Hillmann, M., Hahn, F., Mitschang, B., and Westkämper, E. (2013). The operational process dashboard for manufacturing. In *Procedia CIRP*, volume 7, pages 205–210.
- Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. *Technometrics*, 11(1):1–21.
- Grubbs, F. E. and Beck, G. (1972). Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations. *Technometrics*, 14(4):847–854.
- Hardy, R. J. and Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine*, 17(8):841–856.
- Hart, S. G. and Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, 52(C):139–183.
- Hasson, F., Keeney, S., and McKenna, H. (2000). Research guidelines for the Delphi survey technique. *Journal of Advanced Nursing*, 32(4):1008–1015.
- Hedges, L. V. and Olkin, I. (1982). *Statistical Methods for Meta-Analysis*. Academic press.
- Hicks, B. J. (2007). Lean information management: Understanding and eliminating waste. *International Journal of Information Management*, 27(4):233–249.
- Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., and Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. wiley.
- Higgins, J. P., Thompson, S. G., Deeks, J. J., and Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327(7414):557–560.

BIBLIOGRAPHY

- Hinrichsen, S. and Bendzioch, S. (2019). How digital assistance systems improve work productivity in assembly. In *Advances in Intelligent Systems and Computing*, volume 781, pages 332–342. Springer Verlag.
- Hold, P., Erol, S., Reisinger, G., and Sihm, W. (2017). Planning and Evaluation of Digital Assistance Systems. *Procedia Manufacturing*, 9:143–150.
- Hopkins, L. and Smaill, F. M. (1999). Antibiotic prophylaxis regimens and drugs for cesarean section. In *Cochrane Database of Systematic Reviews*. John Wiley & Sons, Ltd.
- Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., and Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I² Index? *Psychological Methods*, 11(2):193–206.
- Huizenga, H. M., Visser, I., and Dolan, C. V. (2011). Testing overall and moderator effects in random effects meta-regression. *British Journal of Mathematical and Statistical Psychology*, 64(1):1–19.
- Iwańczuk, W. and Guźniczak, P. (2015). Neurophysiological foundations of sleep, arousal, awareness and consciousness phenomena. Part 2.
- Jeffri, N. F. S. and Awang Rambli, D. R. (2021). A review of augmented reality systems and their effects on mental workload and task performance.
- Jesson, J. K., Matheson, L., and Lacey, F. M. (2011). *Doing Your Literature Review: Traditional and Systematic Techniques*. SAGE publications, Inc.
- Johansson, P. E., Malmköld, L., Fast-Berglund, ., and Moestam, L. (2018). Enhancing Future Assembly Information Systems - Putting Theory into Practice. In *Procedia Manufacturing*, volume 17, pages 491–498.
- Karanikas, N., Melis, D. J., and Kourousis, K. I. (2018). The Balance Between Safety and Productivity and its Relationship with Human Factors and Safety Awareness and Communication in Aircraft Manufacturing. *Safety and Health at Work*, 9(3):257–264.
- Kate L. Turabian (2013). *A Manual for Writers of Research Papers, Theses, and Dissertations: Chicago Style for Students and Researchers*. University of Chicago Press.
- Kaushish, J. (2010). *Manufacturing Processes*. PHI Learning Private Limited.
- Keeney, S., Hasson, F., and Mckenna, H. (2010). *The Delphi Technique in Nursing and Health Research*. John Wiley & Sons, Ltd.
- Keller, T., Bayer, C., Bausch, P., and Metternich, J. (2019). Benefit evaluation of digital assistance systems for assembly workstations. In *Procedia CIRP*, volume 81, pages 441–446.
- Kock, N. (2004). The psychobiological model: Towards a new theory of computer-mediated communication based on darwinian evolution. *Organization Science*, 15(3).
- Kock, N. (2005). Media richness or media naturalness? The evolution of our biological communication apparatus and its influence on our behavior toward e-communication tools. *IEEE Transactions on Professional Communication*, 48(2):117–130.
- Kohn, V. and Harborth, D. (2018). Augmented reality - A game changing technology for manufacturing processes? In *26th European Conference on Information Systems: Beyond Digitization - Facets of Socio-Technical Change, ECIS 2018*.

- Krcmar, H. (2015). Informationsmanagement. In *Informationsmanagement*, pages 85–111. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Liang, S. and Roast, C. (2014). Five Features for Modeling Augmented Reality. In *Communications in Computer and Information Science*, volume 434 PART I, pages 607–612. Springer Verlag.
- Lindblom, J. and Thorvald, P. (2014). Towards a framework for reducing cognitive load in manufacturing personnel. *Advances in Cognitive Engineering and Neuroergonomics*, 11:233–244.
- Lipsey, M. W. and Wilson, D. B. (2001). Practical meta-analysis. *SAGE publications, Inc.*
- Lukosch, S., Billingham, M., Alem, L., and Kiyokawa, K. (2015). Collaboration in Augmented Reality. *Computer Supported Cooperative Work: CSCW: An International Journal*, 24(6):515–525.
- Lušić, M., Fischer, C., Bönig, J., Hornfeck, R., and Franke, J. (2016). Worker Information Systems: State of the Art and Guideline for Selection under Consideration of Company Specific Boundary Conditions. In *Procedia CIRP*, volume 41, pages 1113–1118.
- Lyell, D., Magrabi, F., and Coiera, E. (2018). The Effect of Cognitive Load and Task Complexity on Automation Bias in Electronic Prescribing. *Human Factors*, 60(7):1008–1021.
- Masoni, R., Ferrise, F., Bordegoni, M., Gattullo, M., Uva, A. E., Fiorentino, M., Carrabba, E., and Di Donato, M. (2017). Supporting Remote Maintenance in Industry 4.0 through Augmented Reality. *Procedia Manufacturing*, 11:1296–1302.
- Mead, D. and Moseley, L. (2001). The use of the Delphi as a research approach. *Nurse Researcher*, 8(4):4–23.
- Mengist, W., Soromessa, T., and Legese, G. (2020). Ecosystem services research in mountainous regions: A systematic literature review on current knowledge and research gaps.
- Milgram, P. and Kishino, F. (1994). Taxonomy of mixed reality visual displays. *IEICE Transactions on Information and Systems*, E77-D(12):1321–1329.
- Mizell, D. (2020). Boeings Wire Bundle Assembly Project. In *Fundamentals of Wearable Computers and Augmented Reality*, pages 462–482. Lawrence Erlbaum Associates, Inc.
- Mohr, H. (1997). Wissen als Humanressource. In *Humankapital und Wissen*, pages 13–27. Springer Verlag.
- Müllner, M. (2002). *Erfolgreich wissenschaftlich Arbeiten in der Klinik*. Springer Verlag.
- Nelson, J. P. and Kennedy, P. E. (2009). The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment. *Environmental and Resource Economics*, 42(3):345–377.
- Pereira, R. S., Santos, I. C., Oliveira, K. D., and Leão, N. C. (2019). Meta-analysis as a research tool: A systematic review of bibliometric studies in administration. *Revista de Administracao Mackenzie*, 20(5).
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., and Rushton, L. (2006). Comparison of two methods to detect publication bias in meta-analysis. *Journal of the American Medical Association*, 295(6):676–680.

BIBLIOGRAPHY

- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., and Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine*, 26(25):4544–4562.
- Plakas, G., Ponis, S. T., Agalianos, K., Aretoulaki, E., and Gayalis, S. P. (2020). Augmented reality in manufacturing and logistics: Lessons learnt from a real-life industrial application. In *Procedia Manufacturing*, volume 51, pages 1629–1635.
- Powell, C. (2003). The Delphi technique: Myths and realities. *Journal of Advanced Nursing*, 41(4):376–382.
- Sanna, A., Manuri, F., Lamberti, F., Paravati, G., and Pezzolla, P. (2015). Using handheld devices to support augmented reality-based maintenance and assembly tasks. In *2015 IEEE International Conference on Consumer Electronics, ICCE 2015*, pages 178–179.
- Saptari, A., Leau, J. X., and Mohamad, N. A. (2015). The effect of time pressure, working position, component bin position and gender on human error in manual assembly line. In *IEOM 2015 - 5th International Conference on Industrial Engineering and Operations Management, Proceeding*.
- Schmidt, F. L. and Hunter, J. E. (2016). *Methods of Meta-Analysis: Correcting Error and Bias in Research Findings*. SAGE Publications, Inc.
- Sekaran, U. and Bougie, R. (2016). *Research Methods for Business - A skill building approach*. John Wiley & Sons, Ltd.
- Sethi, A. K. and Sethi, S. P. (1990). Flexibility in manufacturing: A survey. *International Journal of Flexible Manufacturing Systems*, 2(4):289–328.
- So, P. (2020). NASA Task Load Index. <https://humansystems.arc.nasa.gov/groups/TLX/>.
- Spath, D., Ganschar, O., Gerlach, S., Hämmerle, M., Krause, T., and Schlund, S. (2013). *Produktionsarbeit der Zukunft-Industrie 4.0*. Fraunhofer Verlag, Stuttgart.
- Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., and Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ (Online)*, 343(7818).
- Stoessel, C., Wiesbeck, M., Stork, S., Zaeh, M. F., and Schuboe, A. (2008). Towards Optimal Worker Assistance: Investigating Cognitive Processes in Manual Assembly. In *Manufacturing Systems and Technologies for the New Frontier*, pages 245–250. Springer London.
- Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R., and Jones, D. R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal*, 320(7249):1574–1577.
- Sutton, A. J. and Higgins, J. P. (2008). Recent developments in meta-analysis. *Statistics in Medicine*, 27(5):625–650.
- Sweller, J. (1988). Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science*, 12(2):257–285.

- Syberfeldt, A., Danielsson, O., and Gustavsson, P. (2017). Augmented Reality Smart Glasses in the Smart Factory: Product Evaluation Guidelines and Review of Available Products.
- Taveggia, T. C. (1974). Resolving research controversy through empirical cumulation: Toward reliable sociological knowledge. *Sociological Methods & Research*, 2(4):395–407.
- Teigen, K. H. (1994). Yerkes-Dodson: A Law for all Seasons. *Theory & Psychology*, 4(4):525–547.
- Terhoeven, J., Schiefelbein, F. P., and Wischniewski, S. (2018). User expectations on smart glasses as work assistance in electronics manufacturing. In *Procedia CIRP*, volume 72, pages 1028–1032.
- Teubner, S., Vernim, S., Dollinger, C., and Reinhart, G. (2018). Worker-centred production management Approaches to flexibility-and productivity-enhancing integration of production workers into an increasingly digitized and connected production system. *ZWF Zeitschrift fuer Wirtschaftlichen Fabrikbetrieb*, 113(10):647–651.
- Töpfer, A. (2010). *Erfolgreich Forschen: Ein Leitfaden für Bachelor-, Master-Studierende und Doktoranden (Springer-Lehrbuch) (German Edition)*. Springer Verlag.
- Vanneste, P., Huang, Y., Park, J. Y., Cornillie, F., Decloedt, B., and Van den Noortgate, W. (2020). Cognitive support for assembly operations by means of augmented reality: an exploratory study. *International Journal of Human Computer Studies*, 143.
- Vernim, S. and Reinhart, G. (2016). Usage Frequency and User-Friendliness of Mobile Devices in Assembly. In *Procedia CIRP*, volume 57, pages 510–515.
- Vernim, S., Wehrle, P., and Reinhart, G. (2016). Entwicklungstendenzen für die Produktionsarbeit von morgen. *ZWF Zeitschrift fuer Wirtschaftlichen Fabrikbetrieb*, 111(9):569–572.
- Viechtbauer, W. and Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2):112–125.
- Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., and Cleven, A. (2009). Reconstructing the giant: On the importance of rigour in documenting the literature search process. In *17th European Conference on Information Systems, ECIS 2009*.
- Webel, S., Bockholt, U., Engelke, T., Gavish, N., Olbrich, M., and Preusche, C. (2013). An augmented reality training platform for assembly and maintenance skills. *Robotics and Autonomous Systems*, 61(4):398–403.
- Wessel, D., Heine, M., Attig, C., and Franke, T. (2020). Affinity for technology interaction and fields of study: Implications for human-centered design of applications for public administration. In *ACM International Conference Proceeding Series*, pages 383–386. Association for Computing Machinery.
- White, G. P. (1996). A meta-analysis model of manufacturing capabilities. *Journal of Operations Management*, 14(4):315–331.
- Willke, H. (2001). *Systematisches Wissensmanagement*. Lucius & Lucius, Stuttgart.
- Wright, H. (1976). The immateriality of information. *Journal of Library History*, 11(4):297–315.
- Wu, J., Feng, M., Liu, Y., Fang, H., and Duan, H. (2019). The relationship between chronic perceived stress and error processing: evidence from event-related potentials. *Scientific Reports*, 9(1).

BIBLIOGRAPHY

Yang, Q., Jiang, D., Sun, J., and Tong, S. (2010). Cortical synchrony change under mental stress due to time pressure. In *Proceedings - 2010 3rd International Conference on Biomedical Engineering and Informatics, BMEI 2010*, volume 5, pages 2004–2007.

A Appendix A - Methodological Approach

A.1 Derivation Literature Data Evaluation

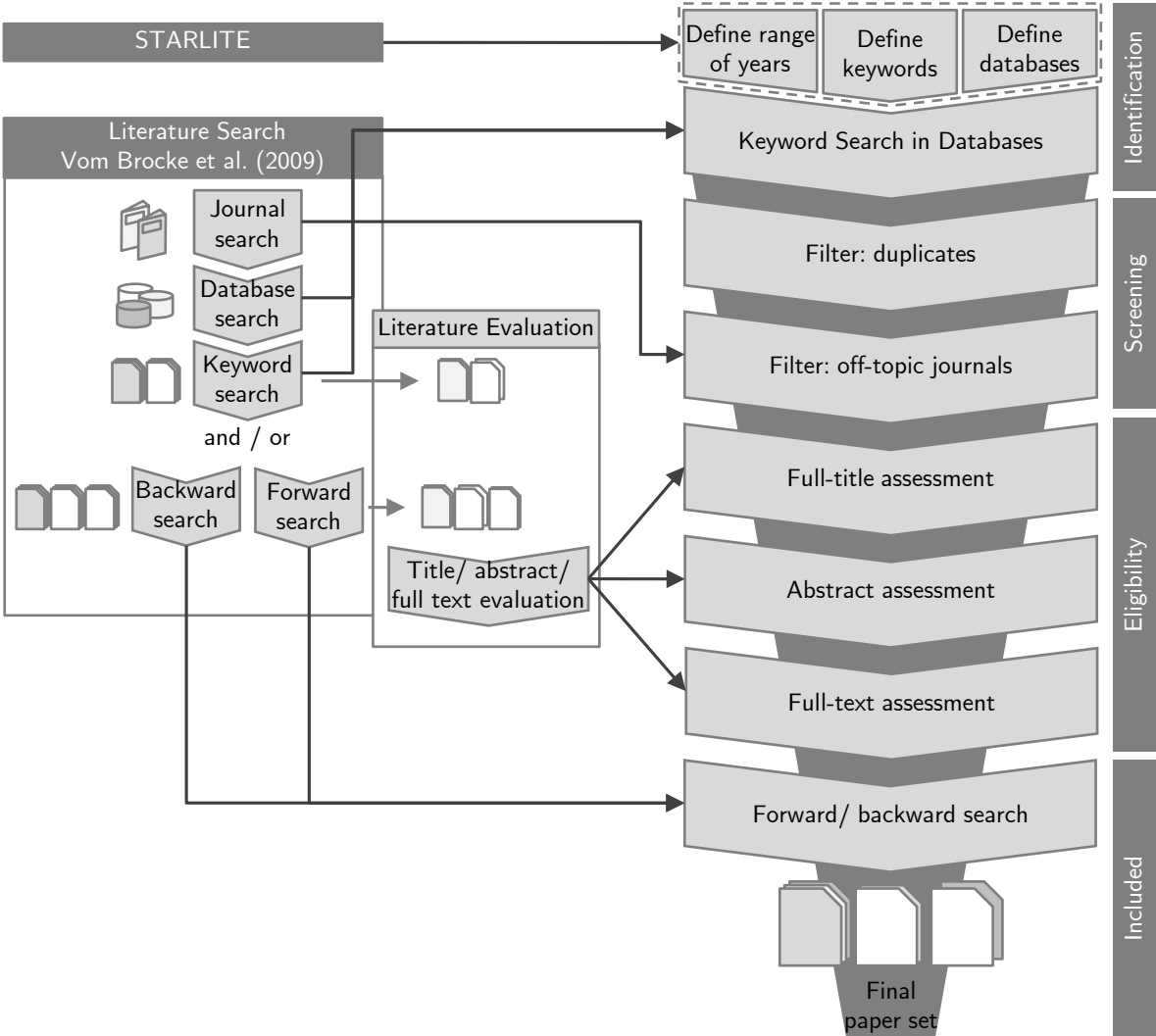


Figure A.1: Derivation evaluation of data following Vom Brocke et al. [2009]

B Appendix B - Quantitative studies

B.1 Overview Quantitative Studies Meta-Analysis

Table B.1: Overview final paper set

#	Author	Titel	Year	Country	Domain	Experimen- tal setting	Classical instruction	Assistance Technology
1	Abbas (1)	Impact of Mobile AR System on Cognitive Behavior and Performance during Rebar Inspection Tasks	2020	Hong Kong	Mainte- nance	Real product/ setting	Paper	AR screen
2	Abbas (2)	Impact of Mobile AR System on Cognitive Behavior and Performance during Rebar Inspection Tasks	2020	Hong Kong	Mainte- nance	Real product/ setting	Paper	HMD
3	Blattgerste (1)	Comparing Conventional and AR Reality Instructions for Manual Assembly Tasks	2017	Germany	Assembly	Lego	Paper	HMD
4	Blattgerste (2)	In-Situ Instructions Exceed Side-by-Side Instructions in Augmented Reality Assisted Assembly	2018	Germany	Assembly	Lego	Paper	HMD
5	Botto	AR for the Manufacturing Industry: The Case of an Assembly Assistant	2020	Italy	Assembly	Mock-up	Paper	AR screen
6	Brice	AugmenTech: The Usability Evaluation of an AR System for Maintenance in Industry	2020	United Kingdom	Mainte- nance	Mock-up	Paper	HMD
7	Büttner	Using Head-Mounted Displays and In-Situ Projection for Assistive Systems – A Comparison	2016	Germany	Assembly	Lego	Paper	HMD
8	Chu	Comparing Augmented Reality-Assisted Assembly Functions – A Case Study on Dougong Structure	2020	China	Assembly	Mock-up	Paper	AR screen
9	Fiorentino	Augmented reality on large screen for interactive maintenance instructions	2014	Italy	Mainte- nance	Mock-up	Paper	AR screen
10	Gavish	Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks	2014	Italy	Mainte- nance + Assembly	Mock-up	Video instruction	HMD
11	Gutsche	Enabling or stressing? – smart information use within industrial service operation	2020	Germany	Mainte- nance + Assembly	Real product/ setting	Paper	HMD
12	Havard	A use case study comparing AR and electronic document-based maintenance instructions considering tasks complexity and operator competency level	2020	France	Mainte- nance	Real product/ setting	PDF tablet	AR screen
13	Hoover	Measuring the performance impact of using the microsoft HoloLens 1 to provide guided assembly work instructions	2020	USA	Assembly	Mock-up	PDF tablet	HMD
14	Hou	Using Animated Augmented Reality to Cognitively Guide Assembly	2014	Australia	Assembly	Lego	Paper	AR screen
15	Lampen	Combining Simulation and Augmented Reality Methods for Enhanced Worker Assistance in Manual Assembly	2019	Germany	Assembly	Mock-up	Paper	HMD
16	Loch	Comparing Video and Augmented Reality Assistance in Manual Assembly	2016	Germany	Assembly	Lego	Video instruction	HMD
17	Obermair	Maintenance with AR Remote Support in Comparison to Paper-Based Instructions: Experiment and Analysis	2020	Austria	Assembly	Real product/ setting	Paper	AR screen

B.1 Overview Quantitative Studies Meta-Analysis

#	Author	Titel	Year	Country	Domain	Experimental setting	Classical instruction	Assistance Technology
18	Pringle	Using an industry-ready AR HMD on a real maintenance task: AR benefits performance on certain task steps more than others	2018	United Kingdom	Maintenance	Mock-up	PDF tablet	AR screen
19	Sanna	Using Handheld Devices to Support Augmented Reality-based Maintenance and Assembly Tasks	2015	Italy	Maintenance + Assembly	Real product/setting	Paper	AR screen
20	Uva	Evaluating the effectiveness of spatial AR in smart manufacturing: a solution for manual working stations	2017	Italy	Assembly	Real product/setting	Paper	AR screen
21	Wang	Usability evaluation of an instructional application based on Google Glass for mobile phone disassembly tasks	2019	Taiwan	Assembly	Real product/setting	Paper	HMD
22	Werrlich	Comparing HMD-based and Paper-based Training	2018	Germany	Assembly	Real product/setting	Paper	HMD
23	Yamaguchi	Video-Annotated Augmented Reality Assembly Tutorials	2020	Austria	Assembly	Real product/setting	Video Instruction	AR screen
24	Yang	Comparing the Effects of Paper and Mobile Augmented Reality Instructions to Guide Assembly Tasks	2020	The Netherlands	Assembly	Lego	Paper	AR screen

Table B.2: Overview data final paper set

#	Author	Number of participants	Ø age	Time ⌚		Error rate ✖		Cognitive load 🧠	
				Classical	Assisted	Classical	Assisted	Classical	Assisted
1	Abbas (1)	15	-	-	-	Ø = 13.51 SE = 5.44	Ø = 9.08 SE = 5.43	-	-
2	Abbas (2)	15	-	-	-	Ø = 13.51 SE = 5.44	Ø = 9.42 SE = 4.29	-	-
3	Blattgerste (1)	24	23.63	Ø = 3.36 SE = 0.49	Ø = 4.3 SE = 0.8	Ø = 1.29 SE = 1.6	Ø = 1.17 SE = 1.43	Ø = 33.13 SE = 17.53	Ø = 40.5 SE = 20.92
4	Blattgerste (2)	24	23.72	Ø = 4.47 SE = 2.27	Ø = 3.88 SE = 1.04	Ø = 0.0182 SE = 0.0275	Ø = 0.0013 SE = 0	Ø = 36.7 SE = 20	Ø = 29.6 SE = 17.4
5	Botto	26	31	Ø = 185 SE = -	Ø = 128 SE = -	Ø = 1.08 SE = -	Ø = 0.15 SE = -	-	-
				p = 0.0016		p = 0.0497			
6	Brice	20	33.5	Ø = 1056 SE = 188	Ø = 1098 SE = 244	-	-	Ø = 39.3 SE = 14.54	Ø = 29.4 SE = 11.95
7	Büttner	13	25.8	Ø = 8.39 SE = 0.73	Ø = 21.09 SE = 6.81	Ø = 0.25 SE = 0.46	Ø = 0.75 SE = 1.75	-	-
8	Chu	16	22.5	Ø = 358.5 SE = 102.48	Ø = 446.38 SE = 81.62	Ø = 4.25 SE = 2.93	Ø = 2.31 SE = 1.92	Ø = 25.8 SE = -	Ø = 32.1 SE = -
								p = 0.16	
9	Fiorentino	14	25	Ø = 13.1 SE = -	Ø = 8.1 SE = -	Ø = 0.094 SE = -	Ø = 0.007 SE = -	-	-
				p < 0.001		p < 0.001			
10	Gavish	10	33.2	Ø = 516 SE = 186	Ø = 492 SE = 120	Ø = 1.3 SE = 1.1	Ø = 0.3 SE = 0.7	-	-
11	Gutsche	10	42.2	Ø = 1337 SE = -	Ø = 949 SE = -	-	-	Ø = 40.67 SE = -	Ø = 35.83 SE = -
				p = 0.199				p = 0.875	
12	Havard	10	22.5	Ø = 1330 SE = -	Ø = 1180 SE = -	-	-	Ø = 56 SE = -	Ø = 48 SE = -
				p = 0.38				p = 0.44	
13	Hoover	35	-	Ø = 7 SE = -	Ø = 1 SE = -	Ø = 1868 SE = -	Ø = 1328 SE = -	-	-
				p < 0.0005		p < 0.0005			
14	Hou	25	-	Ø = 11.91 SE = -	Ø = 7.37 SE = -	Ø = 3.4 SE = -	Ø = 1.3 SE = -	Ø = 13.64 SE = -	Ø = 9.84 SE = -
				p = 0.0001		p = 0.0193		p = 0.0053	
15	Lampen	24	29.25	Ø = 177.76 SE = 33.37	Ø = 167.14 SE = 53.78	Ø = 27.08 SE = 17.81	Ø = 24.17 SE = 10.18	Ø = 47.88 SE = 15.29	Ø = 27.01 SE = 16.71
16	Loch	17	-	Ø = 228 SE = -	Ø = 186 SE = -	-	-	Ø = 53.6 SE = -	Ø = 42 SE = -
				p = 0.127				p = 0.132	
17	Obermair	15	-	Ø = 295.87 SE = 59.2	Ø = 297.07 SE = 66.28	Ø = 0.53 SE = 0.66	Ø = 0.13 SE = 0.35	-	-
18	Pringle	18	22.44	Ø = 108 SE = 33	Ø = 93 SE = 27	Ø = 0.57 SE = 0.23	Ø = 0.38 SE = 0.21	-	-
19	Sanna	6	22	Ø = 671 SE = 172.46	Ø = 630.5 SE = 102.55	Ø = 0.5 SE = 0.55	Ø = 1.33 SE = 0.82	-	-

B.1 Overview Quantitative Studies Meta-Analysis

#	Author	Number of participants	Ø age	Time ⌚		Error rate ✖		Cognitive load 🧠	
				Classical	Assisted	Classical	Assisted	Classical	Assisted
20	Uva	16	24.6	Ø = 610.9 SE = -	Ø = 486.7 SE = -	Ø = 7.03 SE = -	Ø = 1.17 SE = -	-	-
				p < 0.001		p < 0.001			
21	Wang	30	23.77	Ø = 292.63 SE = 127.79	Ø = 316.77 SE = 169.64	Ø = 0.28 SE = 0.65	Ø = 0.15 SE = 0.36	-	-
22	Werrlich	15	18.67	Ø = 600 SE = 121	Ø = 702 SE = 144	Ø = 9.58 SE = 4.127	Ø = 2.98 SE = 1.77	Ø = 48.46 SE = 10.88	Ø = 46.45 SE = 10.47
23	Yamaguchi	16	27.8	Ø = 183 SE = 43	Ø = 182 SE = 51	Ø = 0.6 SE = 1.5	Ø = 0.5 SE = 0.8	Ø = 48.1 SE = 24.3	Ø = 28.4 SE = 22.1
24	Yang	72	24.86	Ø = 313 SE = 142	Ø = 324 SE = 96	Ø = 1.17 SE = 1.63	Ø = 0.5 SE = 0.7	Ø = 19.84 SE = 12.40	Ø = 23.25 SE = 12.26

C Appendix C - Results Meta-Analysis

C.1 Grubbs' test

Following Grubbs [1969], Grubbs' statistic for study i is specified as

$$G_i = \frac{|E_i - \bar{E}|}{S} \tag{C.1}$$

where E_i refers to the effect size of study i , \bar{E} to the mean effect size, and S to the standard deviation. The critical values are extracted from Table of Critical Values for T provided by Grubbs and Beck [1972]. Outliers are only found for the evaluation criterion "error rates". For this evaluation criterion, Grubbs' test was repeated twice.

Time		Error rates		Cognitive load	
Effect sizes	Grubbs' test statistics	Effect sizes	Grubbs' test statistics	Effect sizes	Grubbs' test statistics
-0.262	0.920	0.793	0.301	-0.376	1.374
0.329	0.067	0.812	0.325	0.373	0.054
0.912	1.042	0.078	0.579	0.729	0.734
-0.189	0.799	0.855	0.378	-0.349	1.322
0.085	0.340	0.549	0.002	0.049	0.563
-0.925	2.029 *	-1.685	2.750 **	0.338	0.012
1.360	1.792	0.763	0.265	0.830	0.927
0.147	0.237	1.360	1.000	1.282	1.789 *
0.571	0.473	1.039	0.604	0.518	0.331
0.477	0.315	0.864	0.389	0.186	0.302
0.864	0.963	0.688	0.172	0.827	0.921
1.206	1.534	0.197	0.432	-0.275	1.182
0.233	0.092	0.737	0.232		
0.525	0.395	0.844	0.364	Mean	0.344
-0.019	0.514	-1.097	2.027	Std. dev.	0.524
0.486	0.331	1.257	0.873	# values	12
0.263	0.042	0.244	0.374	Critical value	2.412
1.257	1.619	2.052	1.852		
-0.159	0.748	0.081	0.575		
-0.746	1.731	0.531	0.021		
0.021	0.448	Mean	0.548		
-0.090	0.634	Std. dev.	0.812		
Mean	0.289	# values	20		
Std. dev.	0.598	Critical value	2.708		
# values	22				
Critical value	2.758				

* Furthest from the rest, but not a significant outlier
 ** Significant outlier

Figure C.1: Grubbs' test for evaluation criteria "Time", "Errors", and "Cognitive load"

C Appendix C - Results Meta-Analysis

1st calculation		2nd calculation		3rd calculation	
Error rates		Error rates		Error rates	
Effect sizes	Grubb's test statistics	Effect sizes	Grubb's test statistics	Effect sizes	Grubb's test statistics
0.793	0.301	0.793	0.200	0.793	0.060
0.812	0.325	0.812	0.231	0.812	0.101
0.078	0.579	0.078	0.925	0.078	1.415
0.855	0.378	0.855	0.298	0.855	0.188
0.549	0.002	0.549	0.183	0.549	0.442
-1.685	2.750 **	-	-	-	-
0.763	0.265	0.763	0.154	0.763	0.000
1.360	1.000	1.360	1.092	1.360	1.231
1.039	0.604	1.039	0.587	1.039	0.568
0.864	0.389	0.864	0.313	0.864	0.208
0.688	0.172	0.688	0.035	0.688	0.156
0.197	0.432	0.197	0.737	0.197	1.168
0.737	0.232	0.737	0.112	0.737	0.055
0.844	0.364	0.844	0.280	0.844	0.165
-1.097	2.027	-1.097	2.773 **	-	-
1.257	0.873	1.257	0.929	1.257	1.017
0.244	0.374	0.244	0.663	0.244	1.071
2.052	1.852	2.052	2.180	2.052	2.642 *
0.081	0.575	0.081	0.919	0.081	1.408
0.531	0.021	0.531	0.211	0.531	0.479
Mean	0.548	Mean	0.666	Mean	0.764
Std. dev.	0.812	Std. dev.	0.636	Std. dev.	0.485
# values	20	# values	19	# values	18
Critical value	2.708	Critical value	2.680	Critical value	2.652

* Furthest from the rest, but not a significant outlier

** Significant outlier

Figure C.2: Grubbs' test for evaluation criteria "Errors"

C.2 Results evaluation criterion "time"

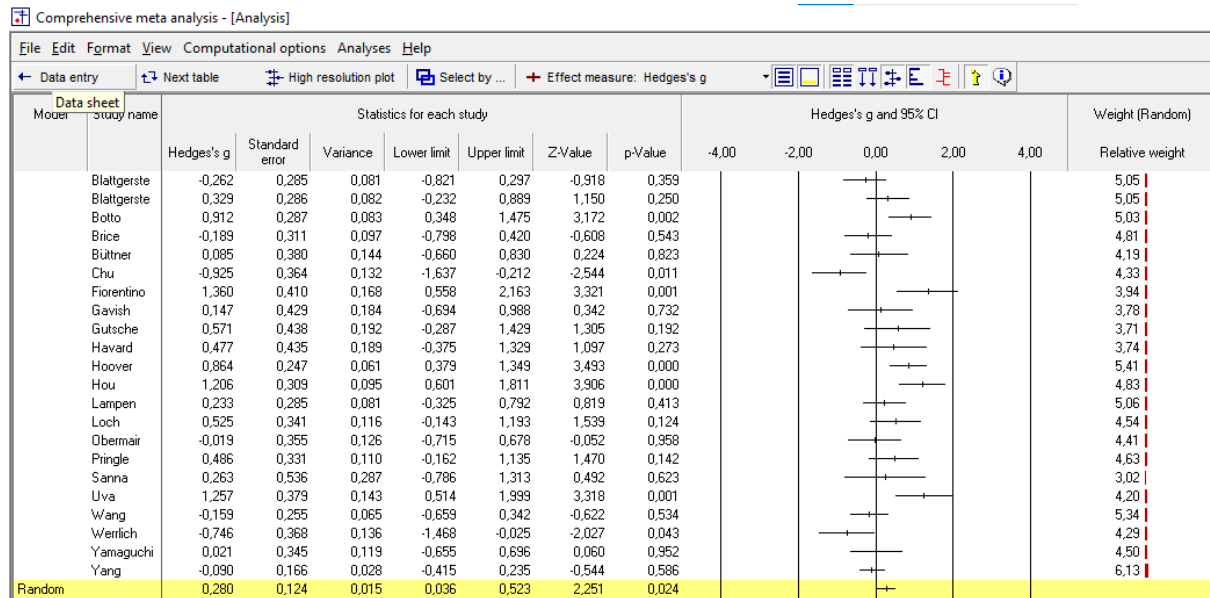


Figure C.3: Results random effects model for evaluation criteria "time" (extracted from CMA)

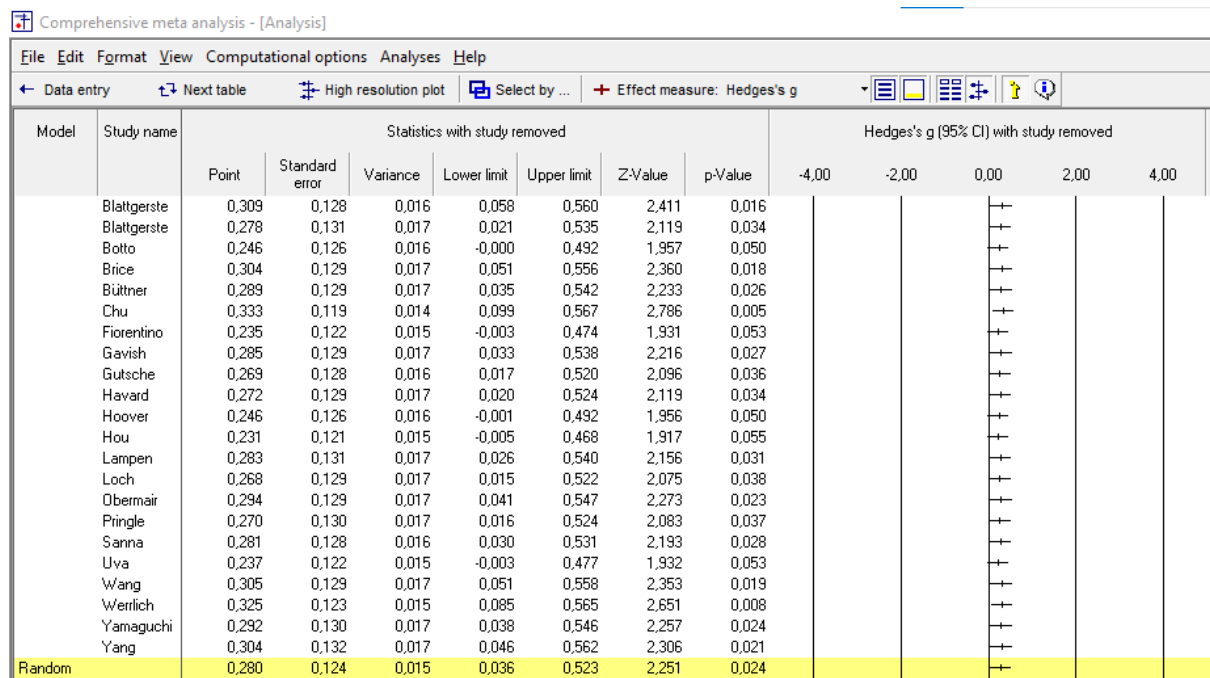


Figure C.4: Results one study removed for evaluation criteria "time" (extracted from CMA)

C.3 Results evaluation criterion "errors"

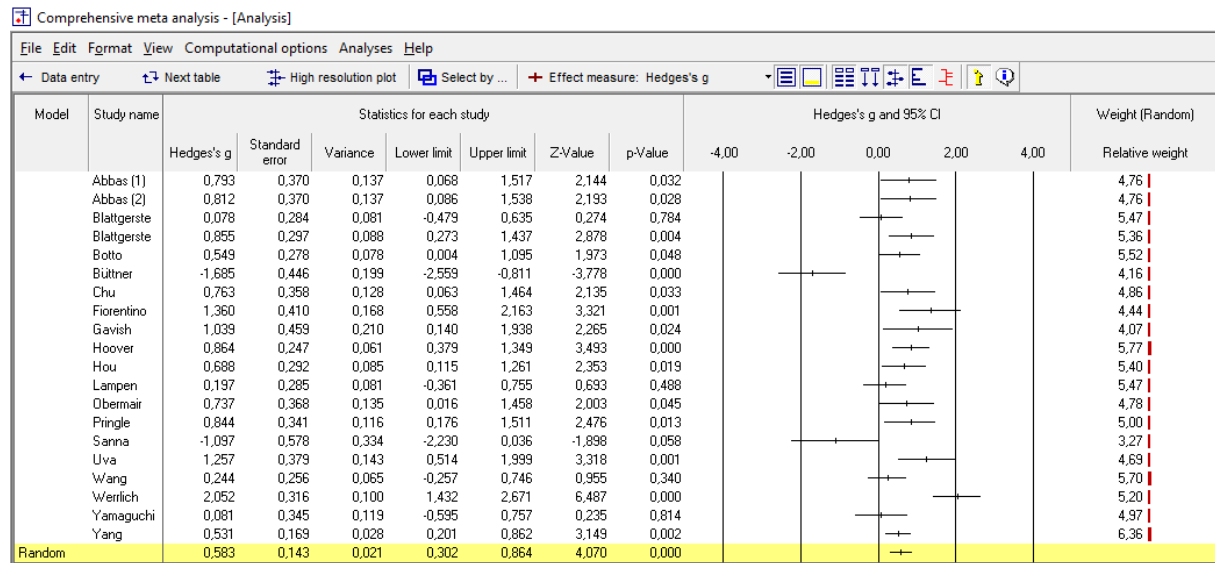


Figure C.5: Results random effects model for evaluation criteria "errors" (extracted from CMA)

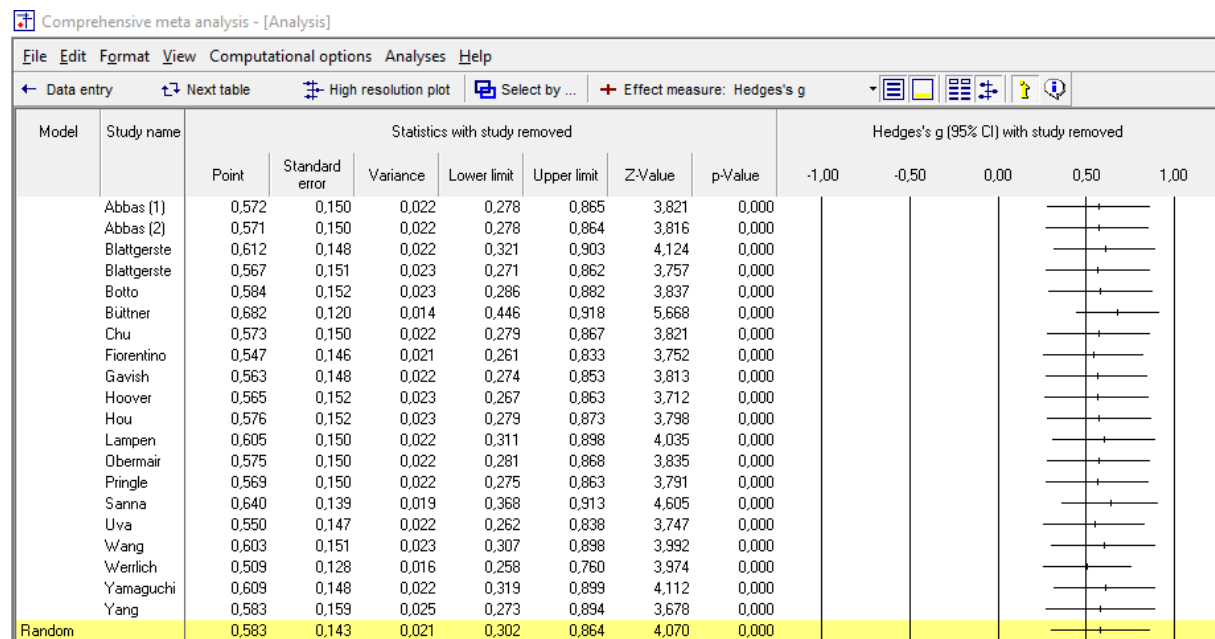


Figure C.6: Results one study removed for evaluation criteria "time" (extracted from CMA)

C.4 Results evaluation criterion "cognitive load"

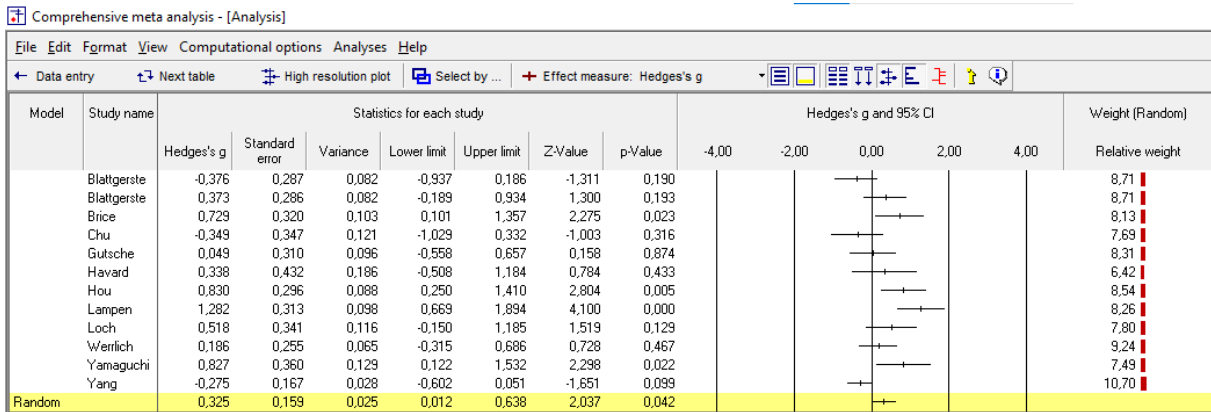


Figure C.7: Results random effects model for evaluation criteria "errors" (extracted from CMA)

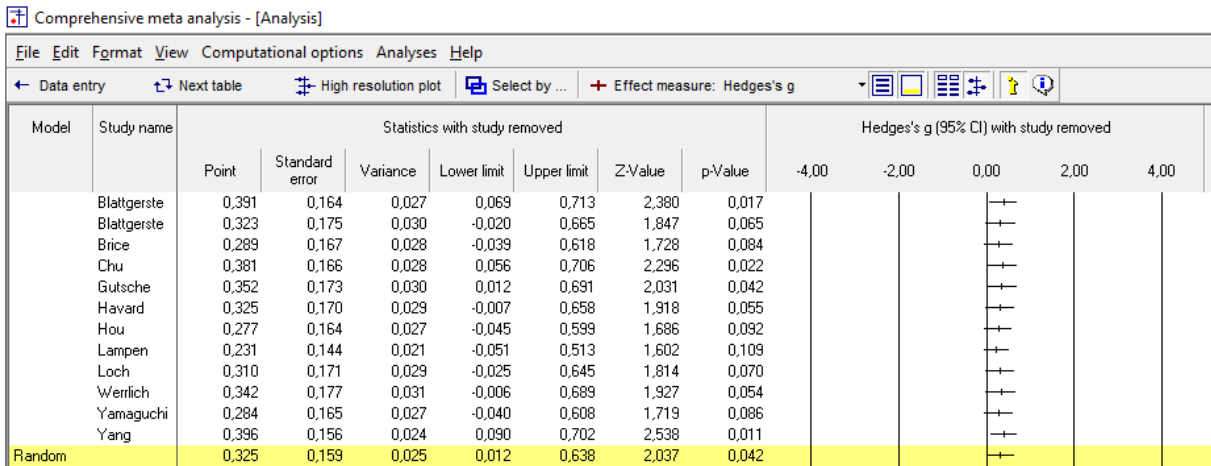


Figure C.8: Results one study removed for evaluation criteria "cognitive load" (extracted from CMA)

C.5 Publication Bias

C.5.1 Duval and Tweedie's Trim and Fill

Table C.1: Results Duval and Tweedie's Trim and Fill (extracted from CMA)

	Time 🕒		Errors ✖		Cognitive load 🧠	
	Observed values	Adjusted values	Observed values	Adjusted values	Observed values	Adjusted values
Studies trimmed	-	0	-	5	-	0
Fixed Effects						
Point Estimate	0.24308	0.24308	0.60195	0.40148	0.22281	0.22281
Lower Limit	0.11275	0.11275	0.46669	0.27849	0.06143	0.06143
Upper Limit	0.37342	0.37342	0.73722	0.52447	0.38419	0.38419
Random Effects						
Point Estimate	0.27980	0.27980	0.58321	0.35513	0.32492	0.32492
Lower Limit	0.03613	0.03613	0.30238	0.05091	0.01233	0.01233
Upper Limit	0.52346	0.52346	0.86404	0.65935	0.63752	0.63752
Q Value	68.52052	68.52052	76.50926	134.28732	38.42477	38.42477

C.5.2 Egger's Regression of the Intercept

Table C.2: Results Egger's Regression of the Intercept (extracted from CMA)

	Time 🕒	Errors ✖	Cognitive load 🧠
Intercept	1.22377	-0.64176	3.88492
Standard error	1.48202	1.71719	1.89590
95% lower limit (2-tailed)	-1.86767	-4.24944	-0.33940
95% upper limit (2-tailed)	4.31521	2.96592	8.10924
t-value	0.82575	0.37373	2.04912
df	20.00000	18.00000	10.00000
P-value (1-tailed)	0.20934	0.35649	0.03380
P-value (2-tailed)	0.41869	0.71297	0.06761

D Appendix D - Results Meta-Regression

As described in Section 6.2, the evaluation criterion "time" is moderated by the country, and the criterion "errors" is moderated by the year. In the following, the attempt to conclude the graphical regressions is explained.

As shown in Figure D.1, low-wage countries such as China and Taiwan show negative effects of AR on the evaluation criterion "time" whereas high-wage countries rather identify positive effects (see, e.g., Australia, France, and the USA). There is no further information available at this point as to why this moderation exists. Furthermore, no meaningful statement can be provided based on the distribution of countries concerning the *Technological Readiness Ranking*. In general, the moderating effects may be related to individual universities and scientists.

At the same time, Figure D.2 highlights that no (linear) trend can be identified in line with the technological developments in recent years. Due to this, a reasonable explanation for the moderating effect is missing at this stage.

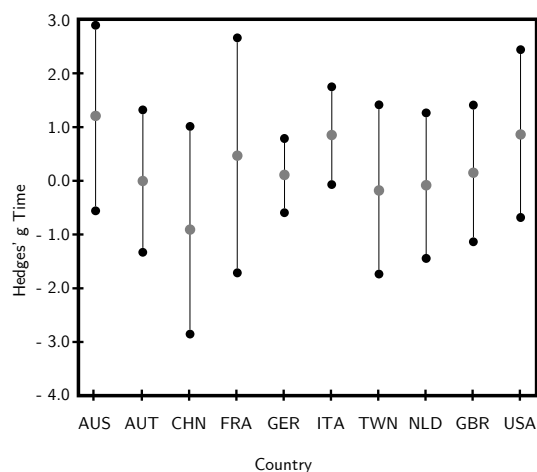


Figure D.1: Regression country on "time"

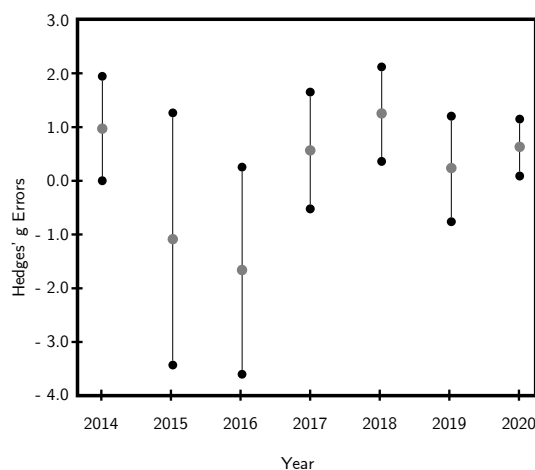


Figure D.2: Regression year on "errors"

E Appendix D - Empirical Exploration

E.1 Results Empirical Exploration

Table E.1: Results individual rankings question related to potentials of AR

Rank	Potential	Points given by participants															Total
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	Faster activities/ processes	5	1	2	2	5	1	7	3	6	1	1	1	3	3	3	44
2	Standardised instructions (same quality output)	1	4	3	3	6	5	4	1	1	3	3	2	1	4	4	45
3	Reduction of errors/ increased reliability	2	3	1	4	3	9	2	9	5	7	2	3	2	1	2	55
4	Independent and flexible instructions	6	2	6	6	1	3	5	2	2	4	7	5	6	5	5	65
5	Increased safety	3	5	5	7	7	7	9	5	4	2	8	4	5	2	1	74
6	Better internal and external communication	4	7	7	1	4	6	1	6	7	6	4	6	4	6	7	76
7	Better problem identification & description	10	10	8	5	2	4	3	10	3	5	6	6	9	10	8	99
8	Hands-free provision of information	8	6	10	8	9	2	6	4	9	8	10	8	8	7	9	112
9	Less cognitive load of users	7	8	4	9	8	8	10	8	10	11	9	7	11	9	6	125
10	Support of home-office	9	11	9	10	10	11	8	7	8	10	5	10	7	8	10	133
11	Social employer branding	11	9	11	11	11	10	11	11	11	9	11	11	10	11	11	159

Table E.2: Results individual rankings question related to challenges of AR

Rank	Potential	Points given by participants															Total
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	Lack of proven business models with positive ROI	2	3	3	1	3	2	5	5	3	6	1	1	4	2	1	42
2	Data protection, privacy, and security	6	9	1	4	2	1	2	2	4	1	4	2	3	4	4	49
3	Technological readiness	1	4	6	4	1	3	4	4	1	4	6	5	2	3	2	50
4	Acceptance from users	3	8	2	5	4	4	3	1	2	5	7	3	1	5	3	56
5	Limited workplace safety and health	7	1	5	2	5	5	1	3	5	10	10	4	6	6	6	76
6	Missing identification of processes/ use cases that are ready to implement AR	5	2	4	6	7	6	7	6	6	7	3	6	5	1	5	76
7	High initial and running costs	4	6	7	9	6	7	6	7	8	9	2	7	8	9	7	102
8	Compatibility with different technologies	10	10	9	10	8	10	9	8	9	2	5	9	7	10	8	124
9	Lack of Development Resources	9	5	8	8	9	8	8	9	7	8	9	10	9	8	9	124
10	Lack of AR App Design & Development Standards	8	7	10	7	10	9	10	10	10	3	8	8	10	7	10	127

Table E.3: Results individual rankings question related to influencing variables of AR

Rank	Potential	Points given by participants															Total
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	Technical affinity of operators and management	1	6	5	4	3	5	8	1	1	1	3	1	2	2	1	44
2	Ease of use of AR applications	5	1	1	3	2	8	7	3	10	3	9	2	1	1	5	61
3	Reliability of technology	2	3	6	7	6	4	4	11	8	5	1	5	4	4	2	72
4	Change management	6	8	3	2	4	3	3	5	5	13	2	6	5	5	4	74
5	Leadership commitment	4	10	4	1	7	6	1	4	3	8	6	3	6	6	6	75
6	Use case readiness	3	2	2	5	12	11	10	8	6	4	4	4	3	3	3	80
7	Degree of digitisation in the manufacturing environment	9	5	10	11	5	7	2	10	2	10	7	10	11	9	9	117
8	Comfort and aesthetics of devices	8	9	9	14	1	1	6	7	11	7	12	8	7	7	10	117
9	Corporate culture incl. curiosity of employees	7	12	7	6	11	12	13	6	7	6	5	7	12	11	12	134
10	Age	12	11	14	9	8	10	9	2	4	2	14	12	10	12	7	136
11	Compatibility with different technologies (AI, IoT, etc.)	13	7	11	8	9	9	5	13	9	9	10	9	9	8	8	137
12	Documentation of instructions and workshops	10	13	8	13	13	2	11	9	12	12	8	11	8	10	13	153
13	In-house app development skills	14	4	12	10	10	13	14	12	13	11	11	13	13	14	14	178
14	Cultural background	11	14	13	12	14	14	12	14	14	14	13	14	14	13	11	197

Table E.4: Results individual rankings question related to use cases of AR

Rank	Potential	Points given by participants															Total
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	Trainings (pre-job) ¹	2	1	3	4	1	5	1	1	2	2	3	1	2	2	1	31
2	Virtual collaboration/ Remote maintenance	1	4	1	1	3	4	2	5	1	1	1	2	3	1	2	32
3	Task guidance (on-the-job) ¹	4	3	2	2	4	1	3	6	3	5	7	4	1	3	4	52
4	Task validation (post-job) ¹	5	2	4	3	7	6	4	2	6	3	4	3	4	4	5	62
5	Virtual factory planning	9	5	5	7	6	7	5	8	4	10	2	8	5	6	6	93
6	Audits	6	10	8	10	5	2	8	4	5	4	5	6	8	8	7	96
7	Plant visits	7	9	9	6	2	3	10	3	7	8	6	9	9	9	9	106
8	Presentation of IoT data	8	7	7	8	10	10	6	9	9	6	10	5	6	5	3	109
9	Navigation in plants	3	8	6	5	9	9	9	10	10	7	8	7	7	7	8	113
10	Material flow visualization	10	6	10	9	8	8	7	7	8	9	9	10	10	10	10	131

¹: individual use cases related to safety, change-over, maintenance tasks, assembly tasks, material handling, machine parameter adjustment, etc.

