

Designing and Embedding Smart Personal Assistant Technology
in Online and Blended Learning Environments

DISSERTATION

of the University of St. Gallen,
School of Management,
Economics, Law, Social Sciences,
International Affairs and Computer Science,
to obtain the title of
Doctor of Philosophy in Management

submitted by

Rainer Winkler

from

Austria

Approved on the application of

Prof. Dr. Jan Marco Leimeister

and

Prof. Dr. Matthias Söllner

Dissertation no. 5160

Niedermann Druck AG, St. Gallen, 2021

The University of St.Gallen, School of Management, Economics, Law, Social Sciences, International Affairs and Computer Science, hereby consents to the printing of the present dissertation, without hereby expressing any opinion on the views herein expressed.

St.Gallen, May 18, 2021

The President:

Prof. Dr. Bernhard Ehrenzeller

To my lovely wife Tanja.

Acknowledgements

This dissertation is the result of three and a half years' work under the supervision of Prof. Dr. Jan Marco Leimeister at the Institute of Information Management at the University of St. Gallen.

First of all, I would like to thank Prof. Dr. Jan Marco Leimeister for his personal support, his supervision of my dissertation, and the stimulating discussions we engaged in, which led to constant critical reflection. Furthermore, I would like to thank Prof. Dr. Matthias Söllner for his willingness to supervise my dissertation, his support in many research projects over the three and a half years, and his relentless efforts to shape me as a researcher and get the best out of me.

Moreover, I would like to thank Susanne Gmünder and Dr. Christina Zenker for their tireless advice, understanding, and support in the course of my PhD. Many others have given me valuable support during my PhD studies. In particular, I would like to thank Dr. Sebastian Hobert and Prof. Dr. Antti Salovaara for our intensive collaboration and endless hours of discussions and the joy of working together.

Furthermore, I am particularly grateful to my co-authors and my colleagues at the Institute of Information Management and the University of Kassel. Working with such a great team has been a constant source of inspiration and motivation during the past years. My special thanks go to Thiemo Wambsganss, with whom I not only published some papers, but also had very fruitful discussions.

I express my special thanks to my mother Anni and my father Hannes Winkler, my sister Marina Köllemann, my parents-in-law Steve and Esther Gähwiler, my American friend Asad Anif, and all my other friends for the encouragement and many methods of support during this trip. In so many ways, I owe my work to them.

Finally, I am deeply grateful to my wife Tanja. It was her love and her enchanting nature that helped me get through the rough patches, always supported me, and provided the perfect grounding for conducting this PhD. She believed in me even during times when I did not believe in myself. For more than 11 years she has not only been my partner but also my very best friend.

St. Gallen, December 2020

Rainer Winkler

List of Publications

During my time as a research assistant and PhD student at the Institute of Information Management in St. Gallen, I wrote and co-authored the following publications under the supervision of Prof. Dr. Jan Marco Leimeister (PhD supervisor). Most of the publications relate to the topic of my dissertation, and consequently they overlap with the contents of this thesis. In each section of the dissertation, when relevant, I provide short notes highlighting publications that have significantly influenced the content of the respective section. The following list of publications is sorted chronologically starting with the most recent (under review, 2020, 2019, etc.)

Under Review

Winkler, R.; Söllner, M.; Leimeister, J. M. (Under Review): Enhancing Problem-Solving Skills with Smart Personal Assistant Technology. In: Computers & Education (C & E).

Elshan, E.; Ebel, P.; Winkler, R.; Söllner, M.; Leimeister, J.M. (Under Review): Because Context Matters - An Integrated Design Approach for Facilitating the Development of Smart Personal Assistants in Low Code Environments. In: Journal of Management Information Systems (JMIS).

Publications in 2020

Winkler, R.; Söllner, M.; Leimeister, J. M. (2020, AOM Global Forum Best Paper): Improving Students' Problem-Solving Skills with Smart Personal Assistants. In: Academy of Management Annual Meeting (AOM), Vancouver, Canada.

Winkler, R.; Hobert, S.; Appius, M. (2020): Empowering Educators to Build their own Conversational Agents in Online Learning. In: European Conference on Information Systems (ECIS), Marrakesh, Morocco.

Winkler, R.; Hobert, S.; Fischer, T.; Salovaara, A., Söllner, M., Leimeister, J. M. (2020): Engaging Students in Online Video Lectures with Dynamically Scaffolding Conversational Agents. In: European Conference on Information Systems (ECIS), Marrakesh, Morocco.

Wambsganss, T.; Winkler, R.; Schmid, P. S.; Söllner, M.; Leimeister, J. M. (2020): Unleashing the Potential of Conversational Agents for Course Evaluations: Empirical Insights from a Comparison with Web Surveys. In: European Conference on Information Systems (ECIS), Marrakesh, Morocco.

Winkler, R.; Briggs, R. O.; De Vreede, G.; Leimeister, J.M.; Oeste-Reiss, S.; Söllner, M. (2020): Modeling Support for Mass Collaboration in Open Innovation Initiatives—The Facilitation Process Model 2.0. In: IEEE Transactions on Engineering Management.

Winkler, R.; Hobert, S.; Salovaara, A.; Söllner, M.; Leimeister, J. M. (2020): Sara, the Lecturer: Improving Learning in Online Education with a Scaffolding-Based Conversational Agent. In: ACM CHI Conference on Human Factors in Computing Systems (CHI), Honolulu, Hawai'i.

Wambsganss, T.; Winkler, R.; Söllner, M.; Leimeister, J. M. (2020): A Conversational Agent to Improve Response Quality in Course Evaluations. In: ACM CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI Extended Abstracts), Honolulu, Hawai'i.

Wambsganss, T.; Winkler, R.; Schmid, P. S.; Söllner, M. (2020): Designing a Conversational Agent as a Formative Course Evaluation Tool. In: Wirtschaftsinformatik (WI), Berlin, Germany.

Winkler, R.; Weingart, P., Söllner, M. (2020): Using Smart Personal Assistants for Online Learning Activities: What benefits can We expect? In: Wirtschaftsinformatik (WI), Berlin, Germany.

Winkler, R.; Söllner, M. (2020): Towards Empowering Educators to Create their own Smart Personal Assistants. In: Hawaii International Conference on System Sciences (HICSS), Maui, Hawaii, USA.

Winkler, R.; Elshan, E.; Söllner, M.; Leimeister, J. M. (2020): SPAM - A Process Model for Developing Smart Personal Assistants. In: Hawaii International Conference on System Sciences (HICSS), Maui, Hawaii, USA.

Publications in 2019

Winkler, R.; Roos, J. (2019): Bringing AI into the Classroom: Designing Smart Personal Assistants as Learning Tutors. In: International Conference on Information Systems (ICIS), Munich, Germany.

Winkler, R.; Büchi, C.; Söllner, M. (2019): Improving Problem-Solving Skills with Smart Personal Assistants: Insights from a Quasi Field Experiment. In: International Conference on Information Systems (ICIS), Munich, Germany.

Winkler, R.; Neuweiler, M. L.; Bittner, E.; Söllner, M. (2019): Hey Alexa, Please Help Us Solve This Problem! How Interactions with Smart Personal Assistants Improve Group Performance. In: International Conference on Information Systems (ICIS), Munich, Germany.

Winkler, R.; Bittner, E.; Söllner, M. (2019): Alexa, Can You Help Me Solve That Problem? Understanding the Value of Smart Personal Assistants as Tutors for Complex Problem Tasks. In: Wirtschaftsinformatik (WI), Berlin, Germany.

Winkler, R.; Söllner, M.; Neuweiler M. L.; Rossini, F. C.; Leimeister, J. M. (2019): Alexa, can you Help us Solve this Problem? How Conversations with Smart Personal Assistant Tutors increase Task Group Outcomes. In: ACM CHI

Conference on Human Factors in Computing Systems Extended Abstracts (CHI Extended Abstracts), Glasgow, Scotland.

Weingart, P.; Winkler, R.; Söllner M. (2019): A SPAT Dialog Model: Understanding human-SPAT Collaboration in a Complex Task Environment. In: Wirtschaftsinformatik (WI), Siegen, Germany.

Winkler, R.; Briggs, R. O.; de Vreede, G.; Leimeister, J. M.; Oeste-Reiß, S.; Söllner, M. (2019): Towards a Technique for Modeling New Forms of Collaborative Work Practices – The Facilitation Process Model 2.0. In: Hawaii International Conference on System Sciences (HICSS), Big Island, Hawaii, USA.

Publications in 2018

Winkler, R.; Söllner, M. (2018): Unleashing the Potential of SPAT in Education: A State-Of-The-Art Analysis. In: Academy of Management Annual Meeting (AOM), Chicago, USA.

Table of Contents

Acknowledgements.....	III
List of Publications.....	IV
Under Review	IV
Publications in 2020	IV
Publications in 2019	V
Publications in 2018	VI
List of Figures	XII
List of Tables	XIII
List of Abbreviations	XVI
Abstract.....	XVIII
Zusammenfassung	XIX
1 Introduction	1
1.1 Problem Statement	1
1.2 Solution Statement and Research Questions.....	3
1.3 Structure of the Dissertation	11
1.4 Overview of Publications in Relation to the Research Questions.....	12
2 Theoretical Background	14
2.1 Technology-mediated Learning	15
2.2 Learning Context: Constructivism.....	17
2.2.1 Implications of Constructivism for Transforming Online and Blended Learning Environments	21
2.2.2 Building Constructivist Learning Environments with Technology	22
2.3 Learning Method Structure: Smart Personal Assistant Technology	24
2.4 Learning Process.....	29
2.5 Learning Outcomes.....	30
3 Philosophical and Methodological Background	32
3.1 Philosophy of Science and Research Epistemology	32
3.1.1 Information Systems Research.....	32
3.1.2 Theoretical Reasoning and Major Streams in Information Systems Research Epistemology	33
3.2 Literature Review	37
3.3 Quantitative Methods.....	38

3.3.1	Experimental Study	38
3.3.2	Quantitative Data Analysis Methods	40
3.4	Qualitative Methods	41
3.4.1	User Interviews	43
3.4.2	Focus Groups	44
3.4.3	Conversation Analysis	44
3.5	NeuroIS Methods	47
3.6	Design Science Research	48
3.6.1	Theory in Design Science Research.....	48
3.6.2	Conducting Design Science Research	53
4	Developing a holistic Understanding of Smart Personal Assistant Technology in Learning Environments.....	56
4.1	Literature Review on the Current State of Smart Personal Assistant Technology in Learning Environments	57
4.1.1	Introduction	57
4.1.2	Research Methodology	58
4.1.3	Results	61
4.1.4	Discussion.....	67
4.1.5	Implications for Research and Practice	69
4.1.6	Limitations.....	71
4.1.7	Conclusion and Next Steps for the Course of the Dissertation....	71
4.2	Requirements of Theory and Practice to Design Smart Personal Assistant Technology	72
4.2.1	Introduction	72
4.2.2	Theoretical Background	73
4.2.3	Research Methodology	73
4.2.4	Designing Smart Personal Assistant Technology in Online and Blended Learning Environments.....	74
4.2.5	Conclusion and Next Steps for the Course of the Dissertation....	80
5	Design of Smart Personal Assistant Technology in Online and Blended Learning Environments	81
5.1	Experimental Evaluation of Smart Personal Assistant Technology to Improve Group Task Outcome	81
5.1.1	Introduction	81
5.1.2	Theoretical Background and Hypotheses Development.....	83

5.1.3	Research Methodology	85
5.1.4	Measurement and Analysis.....	90
5.1.5	Results	94
5.1.6	Discussion.....	97
5.1.7	Limitations and Future Research	99
5.1.8	Conclusion and Next Steps for the Course of the Dissertation..	100
5.2	Using Smart Personal Assistant Technology to increase Long-term Skill Development – Insights from a High School and Vocational School...	101
5.2.1	Introduction	101
5.2.2	Method	101
5.2.3	Experimental Procedure	106
5.2.4	Measurement and Analysis.....	107
5.2.5	Quantitative Results	109
5.2.6	Qualitative Results	111
5.2.7	Discussion.....	112
5.2.8	Limitations and Future Research	115
5.2.9	Conclusion and Next Steps for the Course of the Dissertation..	116
5.3	Experimental Evaluation of Smart Personal Assistant Technology included in an Online Video Lecture to Increase Student Engagement and Outcome	117
5.3.1	Introduction	117
5.3.2	Related Work and Hypotheses Development.....	118
5.3.3	Sara, a dynamically scaffolding SPAT	120
5.3.4	Experimental Evaluation of Sara.....	123
5.3.5	Materials	125
5.3.6	Tests and Questionnaires	126
5.3.7	Procedure	127
5.3.8	Measurement and Data Analysis	127
5.3.9	Results	131
5.3.10	Discussion.....	137
5.3.11	Limitations and Future research	140
5.3.12	Conclusion and Next Steps for the Course of the Dissertation..	141
5.4	Using Smart Personal Assistant Technology for Online Learning Activities – Insights from a Field Study	143
5.4.1	Method	143

5.4.2	SPAT design	143
5.4.3	Sample	145
5.4.4	Experimental Procedure	145
6	Empowering Educators to Design and Embed Smart Personal Assistant Technology in their own Learning Environments	152
6.1	Process Model that Helps Educators to Create and Embed their Own Smart Personal Assistants	152
6.1.1	Introduction	152
6.1.2	End User Development	153
6.1.3	Research Methodology	154
6.1.4	Designing the Process Model	155
6.1.5	Discussion	162
6.1.6	Limitations and Future Outlook	163
6.1.7	Conclusion and Next Steps for the Course of the Dissertation..	163
6.2	SPAT Platform to Empower Educators to Create and Embed Smart Personal Assistants in their own Learning Environments	164
6.2.1	Introduction	164
6.2.2	Theoretical Background	164
6.2.3	Technical Development of MyTeachingBot	166
6.2.4	Summative Evaluation for the SPAT Platform (Step 7)	180
6.2.5	Refinement of Platform and Third Evaluation	184
6.2.6	Contribution	184
7	Summary of Contributions and Areas for Future Research	185
7.1	Theoretical Contributions of the Dissertation	185
7.1.1	Aspects of Smart Personal Assistant Technology that Affect Learning Processes and Learning Outcomes	187
7.1.2	Design of Smart Personal Assistant Technology in Online and Blended Learning Environments	189
7.1.3	Empowering Educators to Create their own Smart Personal Assistant Technology	191
7.1.4	Components of a Design Theory for Providing Individual Support in Online and Blended Learning Environments with the Help of Smart Personal Assistant Technology	192
7.2	Practical Contributions of this Dissertation	197
7.2.1	Design Principles for Educators	197
7.2.2	Smart Personal Assistant Technology Platform	198

7.3 Directions for Future Research	198
7.3.1 Educational Perspective	199
7.3.2 Conversation Design Perspective	201
7.3.3 Technical Perspective	202
8 References	204
Appendix	234
Appendix A Appendices to the study presented in Section 5.2	234
Appendix A.1 Lessons and learning goals	234
Appendix A.2 Homework assignment 2 and characteristics of a problem-based task	234
Appendix A.3 Pre-survey items	235
Appendix A.4 Detailed pre-test results and student characteristics	236
Appendix A.5 Pretest and posttest tasks	237
Appendix A.6 General rating framework per task	238
Appendix A.7 Post-survey items	238
Appendix A.8 Check for assumptions for ANCOVA	239
Appendix B Appendices to the study presented in Section 5.4	240
Appendix B.1 Introductory text and corresponding questions about SWOT analysis	240
Appendix C Appendices to the study presented in Section 6.2	241
Appendix C.1 Route code snippet for creating a SPAT	241
Appendix C.2 Controller code snippet for creating a SPAT	241
Appendix C.3 Model code snippet for creating a SPAT	242
Appendix C.4 Constructs for summative evaluation of SPAT platform	242

List of Figures

Figure 1. Focus of this Dissertation based on TML Framework	4
Figure 2. Process of the Dissertation with Input, Focus and Output Structure..	5
Figure 3. Structure of my Dissertation	11
Figure 4. Technology-mediated Learning Framework.....	15
Figure 5. Learning Scenario Model	23
Figure 6. DSR Knowledge Contribution Framework.....	52
Figure 7. Design Science Research Cycles	53
Figure 8. Design Science Research Process	54
Figure 9. Initial Search Questions derived from our Framework	59
Figure 10. Design Science Research Process of the Paper	73
Figure 11. Requirements and Derivation of DPs	77
Figure 12. SPAT DFs and Excerpt of Example Dialog with SPAT	88
Figure 13. Example Setup with Alexa Echo Dot (left) or scripted human facilitator (right).	89
Figure 14. Experimental Procedure	90
Figure 15. 4 Components Affecting the Collaboration Quality	91
Figure 16. Fragment 1: Asking for Idle Curiosity, TG	95
Figure 17. Fragment 2: Breaks after SPAT Utterance, TG.....	96
Figure 18. Fragment 3: Breaks after Human Facilitator Utterance, Control Group	96
Figure 19. Fragment 4: Laughing after SPA Utterance, TG	97
Figure 20. Example dialogue between the Student and SPAT, and Corresponding Proactive and Reactive logic.	105
Figure 21. Excerpt of additional Script for Homework Assignment 2 provided for Control Group.	106
Figure 22. SPA Experiment Implementation Timeline.....	106
Figure 23. Main- and Sub-dialog Mechanism of SPAT	121
Figure 24. Screenshot of SPAT During an Online Video Lecture, Shown on Top of the Video Screen.	123
Figure 25. Timeline of Video 1 and its Interaction Points.	123
Figure 26. Dynamically Scaffolding vs. Statically Scaffolding SPAT Interaction. Left: Students are Asked for Answers. Right: Students see Hints and can Advance by Clicking or Saying "Next".	124

Figure 27. Experimental Procedure	127
Figure 28. Student #58 GSR and Facial Expression Response for the Dynamically Scaffolding TG.....	130
Figure 29. Normalized Post-test Scores for Information Retention and Transfer Ability for Different Conditions. 50% of the Post-test Scores are within the Error Bars.....	132
Figure 30. Mean Differences in Scaffolding Manipulation	133
Figure 31. Example Dialog Between SPAT and Student	144
Figure 32. Experimental Procedure	146
Figure 33. Design Science Research Process	154
Figure 34. SPAT Process Model for Educators	159
Figure 35. Seven Steps to Create and Test SPAT Platform MyTeachingBot.....	166
Figure 36. MyTeachingBot SPAT Platform.....	167
Figure 37. Initially Created File Directory in Visual Studio Editor	169
Figure 38. MyTeachingBot Architecture Overview	171
Figure 39. Educator's View Architecture	172
Figure 40. Components Directory.....	173
Figure 41. Component with Local State	175
Figure 42. Initializing of Local State.....	175
Figure 43. Link of TextField Input Value to the QuizBot Name Property.....	176
Figure 44. Updating the QuizBot Name Handler	176
Figure 45. Client-Server Communication in Educator's View.....	177
Figure 46. SPAT Editor.....	179
Figure 47. ReteJS Integration into Educator's View.....	179
Figure 48. Predefined Dialog for Conversational Dialog	181
Figure 49. SPAT Future Areas of Research.....	199

List of Tables

Table 1. Overview of Publications in Relation to the Research Questions	14
Table 2. Distinction of Student Behavior according to ICAP Framework	20
Table 3. Taxonomy of Learning Outcomes	31
Table 4. Classification of Scientific Disciplines.....	33
Table 5. Overview of Group Comparison Methods	41

Table 6. Conversation Analysis Approach.....	46
Table 7. Neurophysiological Tools and Focus of Measurement of Neurophysiological Tools	48
Table 8. Design Science Research Contribution Types	49
Table 9. Eight Components of an IS Design Theory	50
Table 10. Components of a DP	51
Table 11. Framework for Defining the Search Scope.....	58
Table 12. Research Gaps structured by the Input-Process-Output Framework	70
Table 13. Sample Characteristics.....	74
Table 14. User-Stories and Requirements from Students and Educators	75
Table 15. Requirements from Theory.....	76
Table 16. DP 1 – Humanness	77
Table 17. DP 2 – Conversational Instruction	78
Table 18. DP 3 – Dynamic Scaffolding.....	78
Table 19. DP 4 – Engaging Dialog Flow.....	79
Table 20. DP 5 – Accessibility	79
Table 21. Sample Characteristics.....	86
Table 22. Group Task and Requirements	87
Table 23. Measurement of Collaboration Quality	92
Table 24. Descriptive Statistics of Treatment and Control Group	94
Table 25. Interrater - Agreement Scores	108
Table 26. Summary of Means	109
Table 27. Summary of Means	110
Table 28. Subtopics Identified from the Focus Group Data	111
Table 29. Sample’s Characteristics	125
Table 30. Summary of the Pretest and Posttest Scores.....	131
Table 31. Average Measurement of Galvanic Skin Response and Facial Expression Score	134
Table 32. Themes Identified for the Open-ended Question Asking Students About their Experience when Interacting with the SPAT.	136
Table 33. Mean and Standard Deviation for the Variables.....	147
Table 34. Regression Results	148
Table 35. Sample Characteristics.....	155
Table 36. User-Stories and Requirements from Students and Educators	157

Table 37. Requirements from Theory	158
Table 38. Ps and Corresponding CRs	162
Table 39. Required Software for Programming MyTeachingBot	168
Table 40. Version Control Program and Cloud Environment	170
Table 41. SPAT Platform Component Overview	171
Table 42. Sample Characteristics	181
Table 43. Quantitative Results of Second Formative Evaluation	182
Table 44. Qualitative Results and Corresponding Change Clusters	184
Table 45. Design Theory for Providing Individual Support in Online and Blended Learning Environments	195

List of Abbreviations

AI	<i>Artificial Intelligence</i>
AN(C)OVA	<i>Analysis Of (Co-)Variance</i>
AOM.....	<i>Academy of Management</i>
ASPIRE.....	<i>Authoring Software Platform for Intelligent Resource in Education</i>
AVS.....	<i>Amazon's Alexa Voice Services</i>
AWS.....	<i>Amazon Web Services</i>
C & E	<i>Computers & Education</i>
CA.....	<i>Conversation Analysis</i>
CHI.....	<i>Conference on Computer Human Interaction</i>
CPS	<i>Collaborative Problem-Solving</i>
CR.....	<i>Change Request</i>
CSCW	<i>Computer-supported Cooperative Work</i>
CTAT	<i>The Cognitive Tutor Authoring Tools</i>
DF	<i>Design Feature</i>
DP	<i>Design Principle</i>
DV	<i>Dependent Variable</i>
ECIS	<i>European Conference on Information Systems</i>
EEG	<i>Electroencephalography</i>
EKG	<i>Electrocardiogram</i>
EUD	<i>End User Development</i>
fEMG.....	<i>facial Electromyography</i>
fMRI	<i>functional Magnetic Resonance Imaging</i>
GIFT.....	<i>Generalized Intelligent Framework for Tutoring</i>
GSR	<i>Galvanic Skin Response</i>
h5.....	<i>Google Scholar h5-index</i>
HCI.....	<i>Human Computer Interaction</i>
HICSS.....	<i>Hawaii International Conference on System Sciences</i>
ICAP	<i>Interactive-Constructive-Active-Passive</i>
ICIS.....	<i>International Conference on Information Systems</i>
IoT.....	<i>Internet of Things</i>
IS	<i>Information Systems</i>

ISA	<i>Intervention Selection Accuracy</i>
IT	<i>Information Technology</i>
ITS	<i>Intelligent Tutoring Systems</i>
JMIS.....	<i>Journal of Management Information Systems</i>
JQ	<i>VHB JOUR-QUAL 3</i>
MAN(C)OVA	<i>Multivariate Analysis Of (Co-) Variance</i>
MEG.....	<i>Magnetoencephalography</i>
MOOC.....	<i>Massive Open Online Course</i>
NLP	<i>Natural Language Processing</i>
P	<i>Pain</i>
PET.....	<i>Positron Emission Tomography</i>
RC.....	<i>Research Challenge</i>
RP	<i>Requirements from Practice</i>
RT	<i>Requirements from Theory</i>
SCR	<i>Skin Conductance Response</i>
SD.....	<i>Standard Deviation</i>
SitPed	<i>Situated Learning Authoring Tool</i>
SPA.....	<i>Smart Personal Assistant</i>
SPAT	<i>Smart Personal Assistant Technology</i>
SPATML	<i>Smart Personal Assistant Mediated Learning</i>
TAM	<i>Technology Acceptance Model</i>
TG.....	<i>Treatment Group</i>
US.....	<i>User Stories</i>
USE	<i>User Story Educator</i>
USS	<i>User Story Student</i>
UX.....	<i>User Experience</i>
WI	<i>Wirtschaftsinformatik</i>
WKWI.....	<i>Wissenschaftliche Kommission für Wirtschaftsinformatik</i>
ZPD.....	<i>Zone of Proximal Development</i>

Abstract

Currently, more and more learning environments exist in whole or in part online. Although these learning environments can be scaled to large audiences, the lack of interaction between educators and students is difficult to replace and remains a primary challenge in the field. This is why Smart Personal Assistant Technology (SPAT) has received attention in the field. SPAT describes computer systems intended to converse with humans. These systems may alleviate the problem of individual support in online and blended learning environments by engaging in natural interaction with the students. Although SPAT seems to be promising for education, three important challenges related to SPAT design emerge and will be tackled in this dissertation. Research Challenge (RC) 1: There is limited knowledge about existing SPAT and corresponding requirements of SPAT in online and blended learning environments. RC2: There is limited knowledge of how to design SPAT to improve students' learning outcomes. RC3: It is not yet clear how non-tech-savvy educators can design and use SPAT on their own to offer individual support.

This dissertation tackles these RCs by following a multi-method approach that identifies research gaps and requirements of SPAT in online and blended learning environments in order to propose design principles (DPs) and design knowledge regarding how educators can use the technology. The dissertation has three main core findings. First, the findings show that there is a lack of consensus on how to design and develop SPAT for online and blended learning environments. Second, based on the first finding, this study proposes five DPs on how to design SPAT in these contexts and investigate the effectiveness of these principles in a series of laboratory and field experiments across different school levels (secondary education, higher education) and school types (high school, vocational business school, university). The findings show that these DPs can have a positive influence on students' learning processes resulting in increased students learning outcomes. It also addresses the third RC by developing a SPAT Platform that empowers non-tech-savvy educators to create their own SPAT in order to make this technology accessible to educators who are not technically inclined. The dissertation contributes to a theory of design and action that explains how non-tech-savvy educators can design effective SPAT in their online and blended learning environments.

Keywords: Smart Personal Assistant Technology; Computer Tutoring Systems; Technology-mediated Learning; Design Science Research.

Zusammenfassung

Aktuell werden in zunehmendem Masse Lernumgebungen eingesetzt, die ganz oder teilweise online organisiert sind. Obwohl diese Lernumgebungen eine hohe Anzahl an Lernenden zulassen, ist die mangelnde Interaktion zwischen Lehrenden und Lernenden nur schwer zu ersetzen und bleibt eine primäre Herausforderung in diesem Setting. Aus diesem Grund hat sich das Interesse an der Technologie rund um Smart Personal Assistants (SPAT) deutlich verstärkt. Mit SPAT werden Computersysteme bezeichnet, die dazu bestimmt sind, mit einem Menschen in natürlicher Sprache zu kommunizieren. Durch diese Systeme kann das Problem der individuellen Unterstützung in Online- und Blended-Learning-Umgebungen reduziert werden, indem sie eine natürliche Interaktion mit den Studierenden ermöglichen. Obwohl SPAT für die Bildung vielversprechend zu sein scheint, sind im Zusammenhang mit dem Design von SPAT drei wesentliche Herausforderungen zu nennen, die im Rahmen dieser Dissertation untersucht werden. Forschungsherausforderung (RC) 1: Es gibt nur begrenztes Wissen über existierende SPAT und die entsprechenden Anforderungen von SPAT in Online- und Blended-Learning-Umgebungen. RC2: Es gibt nur begrenztes Wissen darüber, wie SPAT gestaltet werden kann, damit der Lernerfolg von Studierenden gesteigert wird. RC3: Es ist noch nicht klar, wie nicht technisch versierte Pädagogen SPAT selbstständig gestalten und einsetzen können, um individuelle Unterstützung anzubieten.

Diese Dissertation adressiert die RCs, indem sie einem Multimethodenansatz folgt. Die Untersuchung zielt darauf ab, Forschungslücken und Anforderungen von SPAT in Online- und Blended-Learning-Umgebungen zu identifizieren, um Designprinzipien (DPs) vorzuschlagen und Wissen darüber zu entwickeln, wie Pädagogen die Technologie nutzen können. Im Wesentlichen hat die Dissertation drei Ergebnisse: Erstens zeigt sie, dass es keinen Konsens darüber gibt, wie SPAT für Online- und Blended-Learning-Umgebungen konzipiert und entwickelt werden soll. Zweitens werden auf der Grundlage des ersten Ergebnisses fünf Gestaltungsprinzipien vorgeschlagen, wie SPAT in Lernumgebungen gestaltet werden soll, wobei die Wirksamkeit dieser Prinzipien in einer Reihe von Labor- und Feldexperimenten auf verschiedenen Schulstufen (Sekundarstufe, Hochschule) und in verschiedenen Schultypen (Gymnasium, Berufsschule, Universität) untersucht wird. Anhand der Dissertation lässt sich zeigen, dass die Gestaltungsprinzipien einen positiven Einfluss auf den Lernprozess der Lernenden haben können, was zu besseren Lernergebnissen führt. Die dritte RC wird angesprochen, indem eine SPAT-Plattform präsentiert wird, durch die nicht technisch versierte Pädagogen befähigt werden, ihre eigene SPAT zu erstellen. Die Dissertation stellt einen Beitrag zu einer Theorie des Designs und des Handelns dar, die erklärt, wie nicht technisch versierte Pädagogen SPAT effektiv gestalten und einsetzen können.

Schlüsselwörter: Smart Personal Assistant Technology; Computer-Tutorensysteme; Technologievermitteltes Lernen; Designwissenschaftliche Forschung.

1 Introduction

“Our vision of the future is to have conversational agents on computers guide all phases of instruction, learning, formative assessment, and testing. The student will be holding conversations in natural language with multimedia communication channels, just as they have interact with teachers and mentors for centuries.” (Graesser and McDaniel, 2017)

1.1 Problem Statement

The quotation cited above ventures an outlook on the future potential of SPAT in learning environments. The educational landscape has changed dramatically in recent years. A variety of learning environments have changed from traditional presence events to course units that are undertaken completely or to a large extent online. These online and blended learning environments bring many advantages. Online and blended learning environments enable students to develop effective self-directed learning skills (Kim et al. 2014). With the help of these environments, students are able to identify what they need to learn, find and use online resources, apply the information on the problem at hand, and even evaluate resultant feedback (Kim et al. 2014). Moreover, these learning environments are often far more interactive and memorable than voluminous textbooks or one-sided lectures, and they provide better context, more flexibility for the student, a greater sense of perspective, and more engaging activities than traditional education methods (Dumford and Miller 2018). In addition, the advent of concepts such as massive open online courses (MOOCs) offers new possibilities of reaching a wider audience, resulting in a greater variety of students in terms of age, ethnicity, geography, socio-economic status, and educational background (Christensen et al. 2013). This is why enrollment in online courses has been increasing rapidly since the late 1990s (ICEF 2019).

However, these advantages are accompanied by new challenges. Past research shows that students in online and blended learning courses lack individual support from their educators, resulting in low engagement and high attrition rates (Hone and El Said 2016). In these learning environments, educators find it difficult to offer individual support to all of their students due to financial and organizational restrictions (Rietsche et al. 2018). When students attend these courses, they might have different educational backgrounds, different expectations, and different knowledge gaps, which is why they need individual support. A one-size-fits-all approach leads to low learning outcomes and high attrition rates. It is clear from learning theory that students learn best when they receive individual support consisting of meaningful interactions (Vygotsky 1978).

This emerging challenge is why SPAT provoked interest in the field of information systems (IS) and educational research. SPATs involve computer systems intended to converse with humans in natural language (Hauswald et al. 2015). Prominent examples are Amazon's Alexa, Apple's Siri, Google's Assistant, and Microsoft's Cortana. They can be found on a growing number of devices used in everyday life, and as they are able to hold increasingly complex dialogs, they are being used in more and more environments. Initial studies in this area suggested that this technology would also have positive effects in education (Arend 2018; Graesser and McDaniel 2017). To unleash the full potential of this technology in education, it is necessary to better understand how the technology can support students' individual learning processes in order to address the problem of a lack of individual support in online and blended learning environments. Hence, this thesis identifies three overarching challenges that need to be addressed as outlined below.

Research Challenge 1 – There is limited knowledge about the use and corresponding requirements of Smart Personal Assistant Technology in online and blended learning environments

Currently, only a small number of studies have reported successfully implemented SPAT in learning environments (Arend 2018; Lin et al. 2020). For example, the University of Georgia created SPAT based on IBM's Watson platform called 'Jill Watson', which was developed specifically to handle forum posts by students enrolled in a computer science course (Lip ko 2016). Technology-mediated Learning (TML) as an umbrella term for information technology (IT) in education has often emphasized that input structures, as well as the influence of these structures on the learning process, are crucial for learning outcomes (Gupta and Bostrom 2009). At present, there is no common understanding of how different design aspects of SPAT influence students' individual learning processes and the resulting learning outcomes (Hobert and Wolff 2019). An understanding of these issues is important if design principles for SPAT in education are to be derived. This leads to RC2.

Research Challenge 2 – There is limited knowledge about how to design Smart Personal Assistant Technology to influence learning processes and outcomes in online and blended learning environments

Given that the design and use of SPAT in education is an emerging field, most of the studies in this area are explorative in nature and only few empirical studies exist that rigorously measure the effect of SPAT design on learning processes and outcomes. For example, Dousay and Hall (2018) observed how teachers and administrators perceived the implementation of 90 Amazon's Alexa Echo Dots in four school districts in the US involving approximately 900 students. Most of the existing studies focused on specific implementations of SPAT and failed to provide transferable insights and in-depth evaluations. Little is known about how SPAT can be designed to improve learning processes and lead to higher degrees

of learning success. There has been a failure to look at the issue from an input-output perspective, and consequently it remains unclear how SPAT design aspects influence the internal psychological processes through which students' learning occurs.

Research Challenge 3 – There is limited investigation on how educators can be empowered to create and use Smart Personal Assistant Technology for online and blended learning environments

Despite some proven positive effects of SPAT, the distribution of SPAT in online and blended learning environments is sparse (Nye 2014). If SPAT is implemented at all, it is in tech-savvy online courses where educators have the technological know-how to create SPAT for themselves from scratch (Hooshyar et al. 2016). However, to spread SPAT in other domains too, it is necessary to bridge the gap between the technical knowledge of software designers and the pedagogical knowledge of educators. Educators will not use the technology if it takes them a lot of time to create SPAT or when they have to ask software designers for every change they want to make. Little is known about how non-tech-savvy educators can create and use SPAT effectively for their own learning environments (Sottilare et al. 2016).

1.2 Solution Statement and Research Questions

The objective of this dissertation is to contribute to solving the three outlined challenges. To fulfill this overall objective, the dissertation classifies the RCs and corresponding questions within a framework for online and blended learning environments conceptualized according to the TML framework of Gupta and Bostrom (2009). The framework consists of learning context, learning method structures, and the learning process. As shown in Figure 1, the dissertation focuses on constructivism and online and blended learning environments as the learning context, SPAT as the learning method structure, and the impact of SPAT on the learning process and resulting learning outcomes. The first research question focuses on the use and requirements of SPAT in these learning environments thereby tackling RC1. The second research question focuses on how to design SPAT to improve students' individual learning processes and addresses RC2. The third research question investigates how non-tech-savvy educators can create SPAT for their own learning environments. Although I evaluate SPAT designs in different learning environments with different objectives (e.g., improving problem-solving skills) and resources (e.g., video, text), I do not focus on one specific learning objective or corresponding resource.

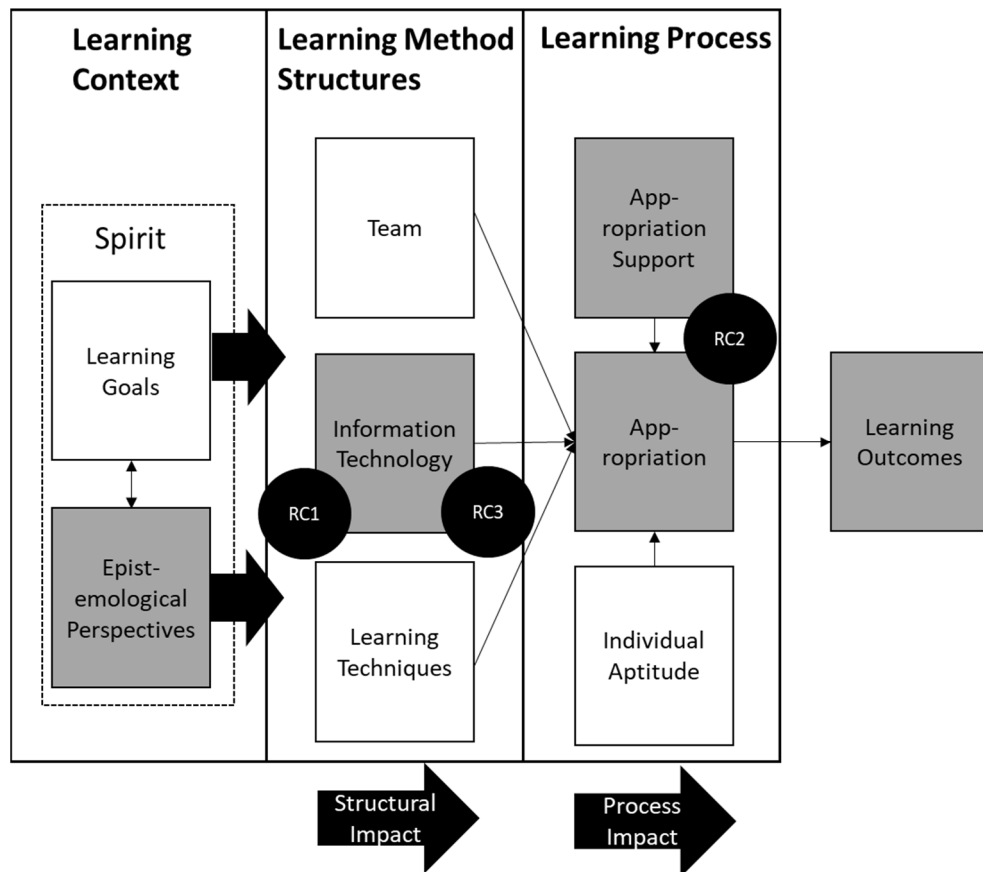


Figure 1. Focus of this Dissertation based on TML Framework

Source: Based on Gupta/Bostrom (2009)

The dissertation focuses on the impact of SPAT on learning processes as a whole. For example, section 5.2 describes how I designed SPAT for problem solving learning activities, where students had to run through structured problem-solving steps to address challenges. I did not focus on single elements of learning activities—unlike, for example, Dadzie et al. (2014). Figure 1 uses the TML framework of Gupta and Bostrom (2009), illustrating the focus of this dissertation (highlighted in dark grey), and highlighting the RCs (displayed as black circles). Based on a certain epistemological perspective (constructivism), I investigate a certain class of IT (SPAT) and how this technology can be used to influence the learning process (appropriation support and appropriation) and resulting metacognitive and cognitive skills (learning outcomes). RC1 relates to the IT itself, whereas RC2 looks at how SPAT influences the learning process. RC3 looks again at the IT and the challenge of empowering educators to create their own SPAT.

Figure 2 depicts the process by which the dissertation tackles the proposed challenges and highlights the input-output relationships between the studies incorporated in this dissertation.

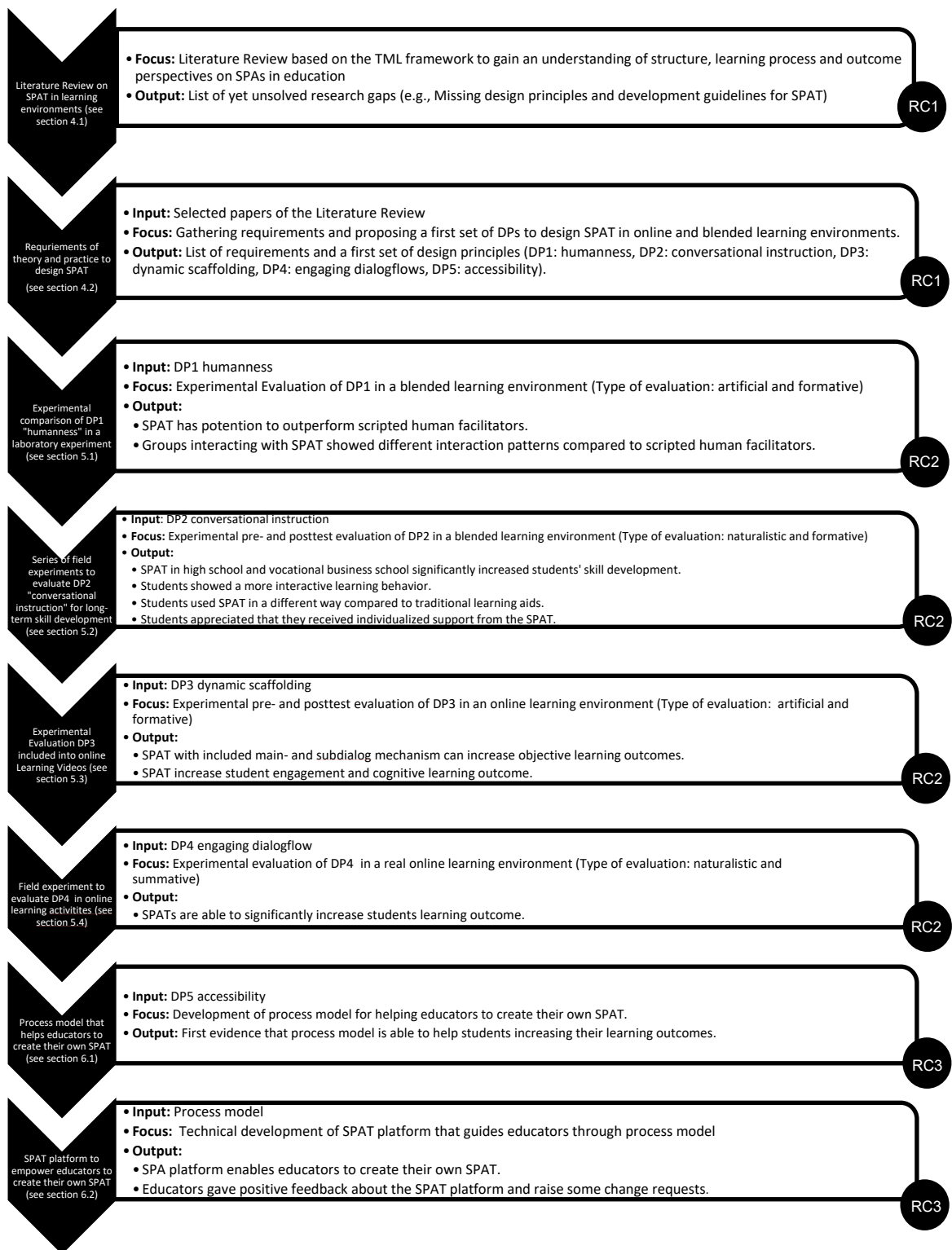


Figure 2. Process of the Dissertation with Input, Focus and Output Structure

Source: Own illustration

RQ1: What are the requirements and relevant design principles of Smart Personal Assistant Technology in online and blended learning environments?

This dissertation begins by describing the current state of the art regarding the use of SPAT in Education and corresponding requirements for SPAT in online and blended learning environments and ends with the definition of design principles. Addressing RQ1 involves dividing it into two parts: First, the literature review considers the current state of the art in SPAT in learning environments from a structure, process, and outcome perspective with the help of the TML model proposed by Gupta and Bostrom (2009). Second, the study derives theoretical and practical requirements for SPAT in online and blended learning environments and formulates a first set of DPs.

RQ1a: What is the state of the art regarding Smart Personal Assistant Technology in learning environments?

Method: Systematic literature review

Results: Structuring the state of the art concerning SPAT research with the help of a TML framework and identifying important research gaps in SPAT research.

The purpose of RQ1a is to gain a holistic understanding of which SPAT exists and how it affects students' learning processes and the resulting learning outcomes. To provide a detailed understanding of SPAT, the disciplines considered in the review process are education, IS, psychology, and management. Altogether, I identified ten research gaps that need more attention. Arising from the structure perspective of the TML research model, there are many unanswered questions on how SPAT influences students' learning processes and the resulting learning outcomes. Moreover, existing studies have shown different ways of evaluating the usefulness of SPAT. From a learning process perspective, further research should be undertaken to investigate the additional value of conversations during the learning process in comparison to often implemented one-directional methods of learning such as tutorial videos in online learning environments. This is an essential step for highlighting the importance of SPAT compared to other traditional learning materials. Moreover, the inclusion of learning theories to identify parts of a learning process that can be supported by SPAT is crucial. From an outcome perspective, more empirical evidence of the influence of SPAT on different kinds of outcomes has to be collected. Until now, most studies have chosen a more explorative research design, failing to measure significant increases on learning outcomes.

RQ1b: What are the relevant design principles of Smart Personal Assistant Technology?

Method: Interview study with 20 students and four educators

Results: Identification of requirements from theory and practice as a basis for a first formulation of design principles

The purpose of RQ1b is to derive requirements from theory and practice in order to translate the requirements into design principles. Among others, the most important requirement of SPAT is the inclusion of a proactive and reactive interaction model, where the SPAT and the student both make a contribution. Moreover, SPAT should provide human-like characteristics with on-the-edge NLP technologies to enable SPAT to ask more complex questions and understand more complicated answers from students. Furthermore, scaffoldings are important in helping students build new knowledge. SPAT should also use an engaging response structure to increase students' learner engagement resulting in higher learning outcomes. SPAT should also allow educators to develop or modify SPAT for themselves rather than being dependent on software designers.

RQ1 addresses RC1 by contributing to a holistic understanding of the current state of SPAT research in education, identifies areas that need more research attention, and serves as a basis for the further testing of design principles.

RQ2: How can Smart Personal Assistant Technology be designed to improve learning processes and learning outcomes in online and blended learning environments?

Building on the results of RQ1 and the corresponding set of DPs, the dissertation evaluates the DPs in online and blended learning environments across different education levels. In RQ2a, I evaluate DP1 (humanness) and DP2 (conversational instruction) in a blended learning environment where students' learning activities are partially home-based and partially class-based. In RQ2b, I switch to an online learning environment to evaluate DP3 (dynamically scaffolding) and DP4 (engaging dialog flow), where students learning activities are entirely outside the class. Among other factors, Interactivity is a fundamental principle for SPAT. The findings of this study suggest that interactivity is essential when both parties need to react to each other and contribute to the dialog in real time. Moreover, SPAT should use human-like attributes to simulate an educator–student discussion. The NLP capabilities of the SPAT are crucial for establishing an educator-like dialog with students. SPAT should provide scaffolds rather than fixed answers to support students' understanding in a meaningful way. Furthermore, SPAT should try to keep the student engaged in the learning process by providing an engaging response structure.

RQ2a: How can humanness and conversational instruction incorporated into Smart Personal Assistant Technology improve students' learning success in blended learning environments?

Method: Mixed method research with a series of experiments and a post-experiment focus group

Results: Empirical evidence for the influence of humanness and conversational instruction incorporated in SPAT on learning outcomes and the influence of it on students learning process in blended learning scenarios.

To test the *humanness* and *conversational instruction* DP, I developed SPAT based on these design principles and tested it first in a blended learning environment. I started with a laboratory experiment, which represents an artificial and formative evaluation according to the FEDS Framework proposed by Venable et al. (2016). Within the lab experiment, groups of three students had to solve a problem task receiving support from SPAT and comparing it with scripted human support and no support. The results revealed that groups interacting with SPAT showed significantly better task outcomes compared to those using scripted human support, and a conversation analysis (CA) approach also identified different patterns in human–SPAT interaction compared to human–human communication. Students asked for hints more frequently, and out of idle curiosity, in the former scenario. Moreover, there were longer gaps between turns and the students laughed more often, which could be an indicator for a more engaged learning process.

Based on these positive findings, I refined DP1 and continued with a natural and summative evaluation (Venable et al. 2016). The study used a mixed-method approach consisting of two field quasi-experiments and one post-experiment focus group discussion at a high school and vocational business school to test DP2 (conversational instruction). Students in the experimental classes completed their assignments using Smart Personal Assistants (SPAs), whereas the control classes completed the same assignments using traditional methods. The empirical results show that students in the experiment classes acquired significantly more problem-solving skills than those in the control classes and also uncovered several changes in students' learning processes. For example, students highlighted the more personalized learning experience when interacting with SPAT, learning at different places than normally, and the positive opportunity to receive individual support when asking something.

In the second part of RQ2, I refined DP1 and DP2 and additionally focused on DP3 and DP4 when creating the new SPAT. Moreover, I tested these DPs in an online learning environment. I started once again with an artificial evaluation in the form of a laboratory experiment and then concluded with an experiment in the field.

RQ2b: How can dynamic scaffolding and an engaging dialog flow incorporated in Smart Personal Assistant Technology improve students' learning success in online learning environments?

Method: Mixed-method research with a series of experiments

Results: Empirical evidence for the influence of dynamic scaffolding and engaging dialog flows incorporated in SPAT on learning outcomes in online learning environments.

The amount of online educational content is rapidly increasing, particularly in the form of online video lectures (Wieling and Hofman 2010). This is why I evaluated

SPAT with DP3 (dynamically scaffolding) layered on top of an existing online video lecture. The SPAT appears during an online video lecture and provides scaffolds by voice and text when needed, and it includes a voice-based input mode. An evaluation with 182 students in a 2 x 2 lab experiment demonstrated that SPAT with DP3 (compared to more traditional SPAT where the design principles were not included) significantly improves learning in a programming task. Moreover, neurophysiological measurements revealed that SPAT significantly improves student engagement (DP4), which is partly explained by differences in students' perceptions of the way they experience the interaction.

After analyzing the findings of this experiment in an artificial setting, I wanted to test DP4 (engaging dialog flow) within parts of real online courses. Hence, I developed SPAT for an online business introductory course learning activity with 76 participants, where one group receives text-based learning materials and the other group receives additional scaffolds from a SPAT. The study shows that students interacting with SPAT are able to achieve higher learning outcomes in a post-test.

After proving the positive effects of the DPs in different kinds of learning outcomes in online and blended learning environments across different domains and school levels, there remains one DP left (DP5, accessibility). To fully incorporate SPAT into online and blended learning environments, it is essential to bridge the gap between the technical knowledge of software designers and the pedagogical knowledge of educators. Educators ultimately decide whether a technology is used in the classroom, and this is why they should be empowered to design and implement SPAT for themselves in an efficient and effective way. Thus, the resulting overarching RQ for the last part of the dissertation is as follows:

RQ3: How can non tech-savvy educators design and embed Smart Personal Assistant Technology in their own online and blended learning environment?

This RQ is divided into two research sub-questions. First, I needed to develop a process model to explore and test how non-tech-savvy educators would design and implement SPAT for their learning environments. Based on that, in RQ3b I developed a SPAT platform that allows educators to run through the different steps of the process model with the help of a SPAT platform.

RQ3a: What are the relevant steps that help non-tech-savvy educators to design and implement SPAT for their own learning environments?

Method: Design science

Results: A process model that shows non-tech-savvy educators how to design SPAT for their own online and blended learning environments.

The objective of this study is to design and validate a process model that describes how educators would design and implement SPAT for their learning environments. I first gathered requirements from educators as well as from IS and

education theory. Next, I created an alpha version of the process model and evaluated it with a focus group and semi-structured interviews. The findings of this study indicate that the process model is able to guide educators in creating SPAT that supports students in online and blended learning environments across different domains and school levels.

RQ3b: How can non-tech-savvy educators be empowered to design and embed SPAT in their own learning environment?

Method: Design science

Results: A SPAT platform that empowers educators to design and implement their own SPAT in order to increase students' learning success.

On the basis of the designed process model developed in response to RQ3a, I gathered requirements from theory on how to design a web interface that empowers educators to design and implement their own SPAT. Next, I created an alpha version of the platform and conducted an evaluation in which 38 non-tech-savvy educators created a SPAT with the platform for both a given environment and their own learning environment. After creating the SPAT, educators completed a questionnaire and I conducted 10 semi-structured post-experiment interviews. The findings reveal that most of the non tech-savvy educators found the platform useful. Moreover, I collected 10 change request (CR) clusters that I incorporated into a beta version.

By formulating these three overarching RQs and answering them in the course of this dissertation, I provide several contributions to SPAT research in particular and IS and education research in general.

The overall objective of this dissertation is to solve the proposed challenges stated in section 1.1, and achieving this is the study's major contribution to research in this area. The dissertation makes three overall theoretical contributions to the IS and education discipline, presenting a theory of explanation and prediction as well as a theory of design and action (Gregor 2006). The theory of explanation and prediction relates to the influencing factors of SPAT that affect students' learning processes and learning outcomes. It helps to explain why SPAT consists of different computer tutoring systems in TML and lays the foundation for the further development of design knowledge. Hence, it contributes to the research area "Influencing factors of computer tutoring systems in TML". Directly connected to this theory, the second main contribution of this dissertation is a theory of design and action, relating to how SPAT has to be designed in order to improve students' learning outcomes in online and blended learning environments. It therefore contributes to the research area "Design principles for computer tutoring systems in TML". The study's final contribution is also a theory of design and action and relates to how non-tech-savvy educators can create SPAT for their own online and blended learning environments. This contributes to the research area "Designing authoring tools for computer tutoring systems in TML".

The dissertation has two major practical contributions. First, it guides software designers in how to design and implement SPAT; and second, it empowers non-tech-savvy educators to create effective SPAT in an efficient way, fostering learning engagement and learning success.

1.3 Structure of the Dissertation

Figure 3 shows how this dissertation addresses the stated goals and RQs.

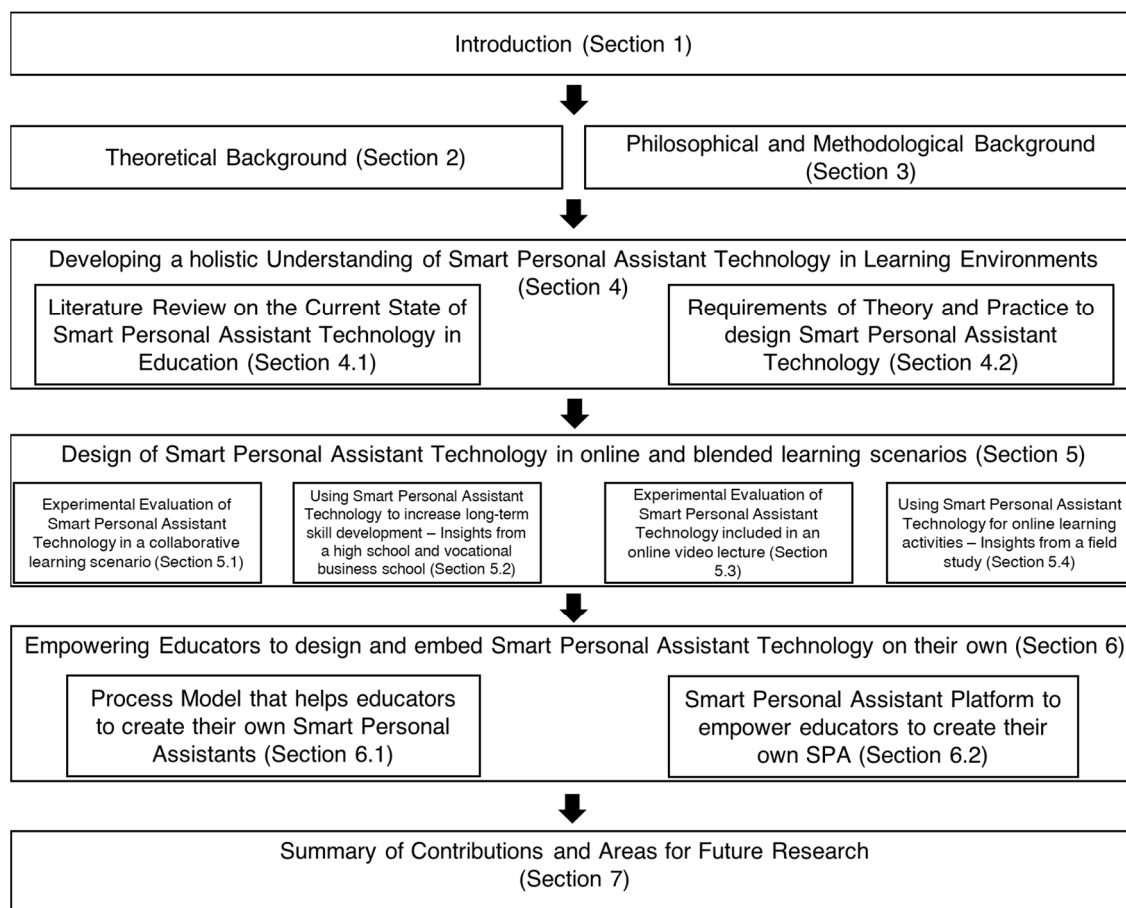


Figure 3. Structure of my Dissertation

Source: Own illustration

In Chapter 2 I provide the necessary theoretical background concerning SPAT research in the area of TML, and Chapter 3 presents the philosophical and methodological background of the dissertation. Based on that, in Chapter 4 the dissertation focuses on developing a holistic understanding of SPAT in learning environments. In Section 4.1 I use a systematic literature review to explore the main research gaps in SPAT research to guide the course of the dissertation and derive requirements. A first set of design principles for SPAT in online and blended learning environments is presented in Section 4.2. In Chapter 5, I evaluate the formulated design principles of SPAT in online and blended learning environments. Starting with a blended learning environment, in Section 5.1 the dissertation investigates the use of two DPs with the help of a laboratory experiment, before relating in Section 5.2 how the DPs were tested in a field experimental

study in a higher school and vocational business school. Based on these findings, in section 5.3 the focus is on an online learning environment, in which I investigate the use of a further two DPs integrated in online videos, first in an online laboratory experiment and then testing the use of SPAT in a real online learning environment. In Chapter 6, I connect to the successfully evaluated design principles by creating a process model and SPAT platform that empowers educators to design and implement their own SPAT. In section 6.1 I describe the development of a process model to describe the process by which educators can design and embed SPAT in their learning environments, and I then describe a web interface that allows non-tech-savvy educators to create and implement SPAT in their online and blended learning environments. Chapter 7 presents a summary of the study's contributions from a theoretical and practical perspective, along with recommendations for future research avenues for SPAT in TML research.

1.4 Overview of Publications in Relation to the Research Questions

The underlying thesis is based on and related to 11 scientific publications that help to answer the outlined RQs. Table 1 contains a list of publications that influenced the content of this dissertation with additional information regarding outlet metrics (if available) such as the Google Scholar h5-index (h5), the outlet rankings of VHB JOURQUAL 3 (JQ3), and the WI-Journal list 2009 of the "Wissenschaftliche Kommission für Wirtschaftsinformatik (WKWI). In addition, the Table illustrates the predominant methodological focus as well as the research context of the publications. At the beginning of each chapter, I also highlight how these publications and the corresponding presentations have influenced the content of each section.

No.	Publication	Outlet Metrics	Method and Context	RQ
1	Winkler, R.; Söllner, M. (2018): Unleashing the Potential of SPAT in Education: A State-Of-The-Art Analysis. In: Academy of Management Annual Meeting (AOM), Chicago, USA.	H5: 3 JQ3: / WKWI: /	Literature review; Deductive; Technology-mediated learning	1a
2	Winkler, R.; Roos, J. (2019): Bringing AI into the Classroom: Designing Smart Personal Assistants as Learning Tutors. In: International Conference on Information Systems (ICIS), Munich, Germany.	H5: 30 JQ3: A WKWI: A	Interview study; Inductive; Higher education and high school, online and blended learning environment	1b
3	Winkler, R.; Söllner, M.; Neuweiler M. L.; Rossini, F. C.; Leimeister, J.	H5: 95 JQ3: /	Quantitative; Deductive;	2a

	M. (2019): Alexa, can you help us solve this problem? How conversations with smart personal assistant tutors increase task group outcomes. In: ACM CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI Extended Abstracts), Glasgow, Scotland.	WKWI: /	Higher education, blended learning environment	
4	Winkler, R.; Neuweiler, M. L.; Bittner, E.; Söllner, M. (2019): Hey Alexa, Please Help Us Solve This Problem! How Interactions with Smart Personal Assistants Improve Group Performance. In: International Conference on Information Systems (ICIS), Munich, Germany.	H5: 30 JQ3: A WKWI: A	Mixed-Method (Quantitative and Qualitative); Deductive and Inductive; Higher Education, blended learning environment	2a
5	Winkler, R.; Büchi, C.; Söllner, M. (2019): Improving Problem-Solving Skills with Smart Personal Assistants: Insights from a Quasi Field Experiment. In: International Conference on Information Systems (ICIS), Munich, Germany.	H5: 30 JQ3: A WKWI: A	Mixed-Method (Quantitative and Qualitative); Deductive and Inductive; High school and vocational business school, blended learning environment	2a
6	Winkler, R.; Söllner, M.; Leimeister, J. M. (Under Review): Enhancing Problem-Solving Skills with Smart Personal Assistant Technology. In: Computers & Education <i>Global Forum Best Paper Winner of the MED Division AoM 2020</i>	H5: 101 JQ3: / WKWI: /	Mixed-Method (Quantitative and Qualitative); Deductive and Inductive; High school and vocational business school, blended learning environment	2a
7	Winkler, R.; Hobert, S.; Salovaara, A.; Söllner, M.; Leimeister, J. M. (2020): Sara, the Lecturer: Improving Learning in Online Education with a Scaffolding-Based Conversational Agent. ACM CHI Conference	H5: 95 JQ3: / WKWI: /	Design Science; Deductive; Higher education, online learning environment	2b

	on Human Factors in Computing Systems (CHI), Honolulu, Hawai'i.			
8	Winkler, R.; Hobert, S.; Fischer, T.; Salovaara, A., Söllner, M., Leimeister, J. M. (2020): Engaging Students in Online Video Lectures with Dynamically Scaffolding Conversational Agents. In: European Conference on Information Systems (ECIS), Marrakesh, Morocco.	H5: 34 JQ3: B WKWI: A	NeuroIS; Deductive; Higher education, online learning environment	2b
9	Winkler, R.; Weingart, P., Söllner, M. (2020): Using Smart Personal Assistants for Online Learning Activities: What benefits can I expect? In: Wirtschaftsinformatik (WI), Berlin, Germany.	H5: 23 JQ3: C WKWI: A	Quantitative; Deductive; Higher education, online learning environment	2b
10	Winkler, R.; Söllner, M. (2020): Towards Empowering Educators to Create their own Smart Personal Assistants. In: Hawaii International Conference on System Sciences (HICSS), Maui, Hawaii, USA.	H5: 30 JQ3: C WKWI: A	Design Science; Deductive; Higher Education, high school and secondary school, online and blended learning environment	3a
11	Winkler, R.; Hobert, S.; Appius, M. (2020): Empowering Educators to Build their own Conversational Agents in Online Learning. In: European Conference on Information Systems (ECIS), Marrakesh, Morocco.	H5: 34 JQ3: B WKWI: A	Design Science; Deductive; Higher Education, high school and secondary school, online and blended learning environment	3b

Table 1. Overview of Publications in Relation to the Research Questions

Source: Own illustration

2 Theoretical Background

“One can design an aeroplane wing on the basis of tested, technological (black-box) rules, but such wings can be designed

*much more efficiently on the basis of tested **and** grounded technological rules, grounded on the laws and insights of aerodynamic and mechanics.” (van Aken 2004, p. 224)*

2.1 Technology-mediated Learning

To show how the outlined challenges that I presented in the previous section are linked together, I refer to the theoretical framework of Gupta and Bostrom (2009). The framework considers structural and outcome factors of TML, and in contrast to previous research approaches it also considers the student learning process, which is crucial for the resulting learning outcomes and thus for the quality of TML (Gupta and Bostrom 2009). This is in line with the predominant constructivist principles of learning that put the students at the center of learning. The framework is displayed in Figure 4.

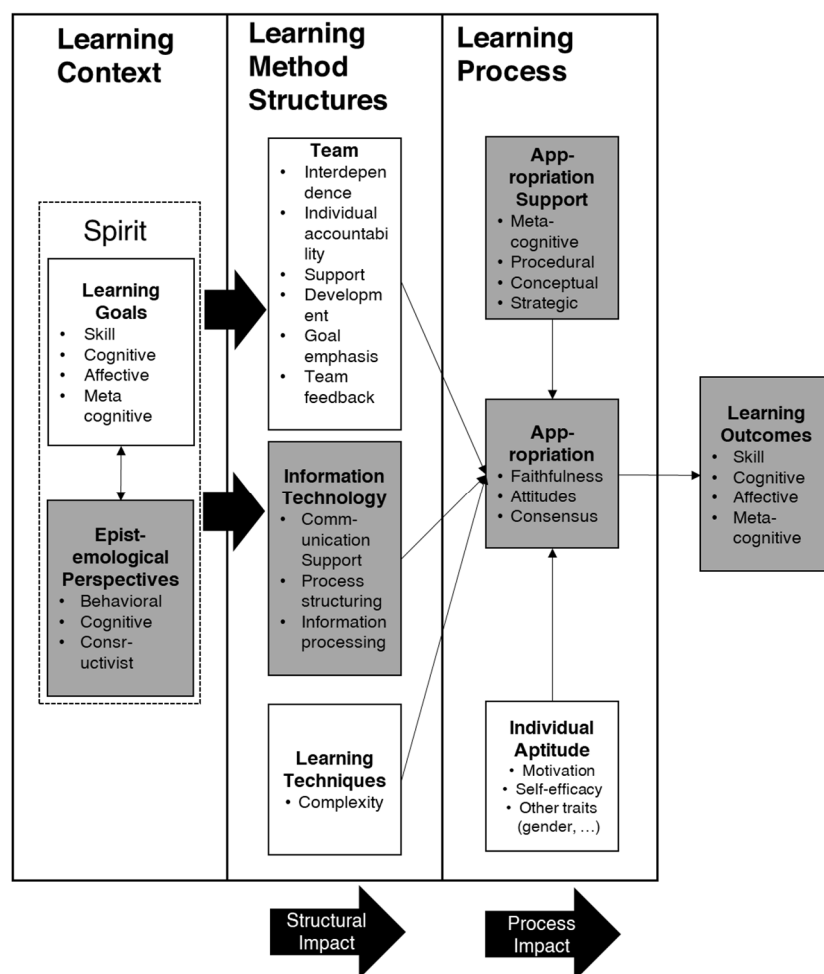


Figure 4. Technology-mediated Learning Framework

Source: Based on Gupta/Bostrom (2009)

The suggested framework is based on two premises. The first relates to the influence of structures embedded in a specific context (DeSanctis and Poole 1994). The context in this dissertation corresponds to a constructivist understanding of learning, and SPAT represents the used IT as a learning method structure (e.g.,

in high schools, higher education, etc.). The second relates to the learning process. The learning process describes how students interact with the previously described learning method structures (SPAT) and how students adapt them (Gupta and Bostrom 2009). The learning process in itself is a complex phenomenon and includes cognitive processes and interactions relating to the already introduced learning methods, individual differences between students, support in the learning process (e.g., with SPAT), and other elements in the learning environment influencing learning outcomes. Learning outcomes represent “the goal assessment or measures for determining the accomplishment of learning goals” (Gupta and Bostrom 2009, p. 691) and are the key outcome measure of TML.

This dissertation is based on a constructivist view of learning that states that it is achieved by students “constructing” knowledge from their experiences (Glaserfeld 1987). In particular, I am interested in how SPAT as a new class of IT can influence students’ learning processes and resulting learning outcomes. Reflecting this, Figure 4 shows the dissertation’s areas of focus, which are learning context, IT, appropriation, and learning outcome clusters. The thesis is less concerned with the team, the complexity of the learning techniques, and the individual characteristics of the students.

Epistemological perspectives are essential for the TML environment design. Epistemology in general “establishes overarching beliefs about the nature of knowledge and about what it means to “know” something (Hannafin et al. 2004). The most common differentiation in education and IS research is between behaviorist, cognitivist and constructivist views of learning (Leidner and Jarvenpaa 1995). All three learning perspectives are based on central assumptions. When considering behaviorism, the central theoretical assumption is that human behavior is predictable and learning occurs when students show new behavior or change their behavior in response to instructional stimuli (Gupta and Bostrom 2009). For example, frontal teaching without interaction is a prototypic learning method grounded in the behavioristic perspective (Ertmer and Newby 1993). When considering the cognitivist view, it is determined by the theoretical assumption that the cognitive structures in student brains as they process information are essential for the learning process (Ertmer and Newby 1993). Learning takes place when students’ knowledge is changed through modified mental representations (or experience) (Mayer and Alexander 2011). In contrast to the behavioristic view, the student has more control over learning. For example, the cognitive process of the provided information to accomplish the task is left to the student and is not prompted by direct instruction. The behaviorist and cognitivist perspectives have in common the idea that the change of knowledge (cognitive change) has to be inferred by behavior (behavioral change) from outside (Ertmer and Newby 1993). This essentially distinguishes it from the constructivist view of learning.

This dissertation focusses on the constructivist view of learning, which is explained in more detail in the next section.

2.2 Learning Context: Constructivism

The constructivist perspective describes learning as a change in meaning constructed from experience (Glaserfeld 1987). Constructivists believe that knowledge and truth are constructed by people and do not exist outside the human mind (Jonassen 1991). In contrast to that, behaviorists and cognitivists believe that knowledge and truth exist outside the human mind and are therefore objective (Runes 1984). From a behaviorist and cognitivist perspective, the way to acquire knowledge is to tell students about the world and expect them to replicate (behaviorist view) or process (cognitivist view) its content and structure in their thinking (Jonassen 1991). It is asserted that there is a particular body of knowledge that needs to be transmitted to students, and learning corresponds to the acquisition and accumulation of a finite set of skills and facts.

From a constructivist perspective, learning was defined by von Glaserfeld (1987) in these terms:

... students construct understanding. They do not simply mirror or reflect what they are told or what they read. Students look for meaning and will try to find regularity and order in the events of the world even in the absence of full or complete information.

Constructivism emphasizes the construction of knowledge, while behaviorism and cognitivism are mainly concerned with the object of knowing. These fundamentally different perspectives on knowledge and learning have major implications for the design of instruction (Tam 2000). The main argument of constructivism is that learning must occur in an active process where information may be imposed, but understanding and the whole process of knowledge creation must come from within. Woolfolk (2000) describes the learning process as follows:

... The key idea is that students actively construct their own knowledge: the mind of the student mediates input from the outside world to determine what the student will learn. Learning itself is active mental work, not passive reception of teaching.

During this active mental work, students may perceive the external reality differently, based on their unique experiences of the world and their beliefs about them (Jonassen 1991). For example, students may discuss their understandings with others helping them to align their understandings to a shared understanding (Cognition and Technology Group 1991). This does not mean that every understanding is legitimate and correct (Spiro et al. 1991). Students must be able to justify their position to establish its viability (Cognition and Technology Group 1991).

According to the constructivist perspective, learning is determined by the complex interplay between students' existing knowledge, the social context, and the problem to be solved. Instruction refers to providing students with a collaborative situation in which they have both the means and the opportunity to construct "new

and situationally-specific understandings by assembling prior knowledge from diverse sources” (Ertmer and Newby 1993). Educators thus slip from the role of the knowledge provider into the role of the coach and preparer for learning. Two characteristics seem to be central for constructivist learning processes. First, constructivist instruction asks students to use their knowledge to solve problems that are meaningful and realistically complex rather than presenting final solutions, and it only offers scaffolding when students require it. The problems provide the context for the students to apply their knowledge and take ownership of their learning. Good problems are required to stimulate the exploration and reflection necessary for knowledge construction. Second, the constructivist perspective supports the notion that students learn through interaction with others. Students work together as peers, applying their combined knowledge to the solution of the problem. The dialogue that results from this combined effort provides students with the opportunity to test and refine their understanding in an ongoing process leading to better learning outcomes.

According to Copley et al. (1992), constructivist learning environments require educators who act as facilitators

... whose main function is to help students become active participants in their learning and make meaningful connections between prior knowledge, new knowledge, and the processes involved in learning.

Chaiklin (2003) stated that educators can best encourage students’ learning by presenting tasks that they can complete only with assistance, meaning that they are in their own *Zone of Proximal Development (ZPD)*. Vygotsky (1978) defined the ZPD as the discrepancy between a student’s actual mental state and the level they reach when they solve problems with assistance. This means that when an educator creates challenging tasks and offers the right amount and type of individual support, students have the most effective learning process and corresponding learning success.

Chung (1991) characterized constructivist learning environments and proposed these four criteria: a shared knowledge among educators and students; a shared authority and responsibility among educators and students; the educator’s new role as guide in instruction; and heterogeneous and small groupings of students. Moreover, within these learning environments, Chung (1991) proposed the following main activities to be carried out by an educator:

- encourage and accept student autonomy and initiative
- use a wide variety of materials
- inquire about students’ understandings of concepts before sharing the educator’s own understanding of those concepts
- encourage students to engage in dialogue with the educator and other students

- encourage student inquiry by asking thoughtful, open-ended questions
- encourage students to ask each other questions
- engage students in experiences that show contradictions to initial understandings
- provide time for students to construct relationships
- assess students' understanding through application and performance of open-structured tasks.

Researchers and educators in the constructivist learning domain have long recognized that although students can learn from receiving information passively, they learn much better by learning actively (Bonwell and Eison 1991). Learning actively requires students to engage cognitively and meaningfully with the tasks they are dealing with. They think about their learning material in depth rather than just passively receiving it (King 1993). Chi and Wylie (2014) called this student behavior an “interactive behavior” and developed a framework that breaks down observable student behaviors into four modes: Interactive, Constructive, Active, and Passive (ICAP-framework). It predicts that these modes will be ordered by effectiveness as follows: Interactive > constructive > active > passive. Each mode of the ICAP framework corresponds to different types of behaviors and knowledge-change processes predicting different learning outcomes. Table 2 outlines the different modes (from passive to interactive), showing that for different learning activities (listening to a lecture, reading a text, and observing a video) students can show different modes of behavior that affect their learning process and success.

	PASSIVE (Receiving)	ACTIVE (Manipulating)	CONSTRUCTIVE (Generating)	INTER- ACTIVE (Dialoguing)
Effectiveness	Not effective	Less effective	More effective	Most effective
Listening to a lecture	Listening without doing anything else but oriented toward instruction	Repeating or rehearsing; Copying solution steps; Taking verbatim notes	Reflecting out loud; Drawing concept maps; Asking questions	Defending and arguing a position in dyads or small group
Reading a text	Reading entire text passages silently/aloud without doing anything else	Underlining or highlighting; Summarizing by copy- and delete	Self-explaining; Integrating across texts; Taking notes in one's own words	Asking and answering comprehension questions with a partner

Observing a video	Watching the video without doing anything else	Manipulating the tape by pause, play, fast-forward, rewind	Explaining concepts in the video; Comparing and contrasting to prior knowledge or other materials	Debating with a peer about the justifications; Discussing similarities & differences
--------------------------	--	--	---	--

Table 2. Distinction of Student Behavior according to ICAP Framework

Source: Adapted from Chi and Wylie (2014)

A *passive* student behavior describes behavior in students who deal with the presented instructional information without additional physical activity (e.g., reading a text without doing anything else). An *active* student behavior involves “doing something physically” (e.g., underlining or highlighting a text). A *constructive* student behavior requires “producing outputs that contain ideas that go beyond the presented information” (e.g., taking notes in one’s own words). An *interactive* behavior requires “dialoguing extensively on the same topic, and not ignoring a partner’s contribution”. This means that both partners’ utterances are primarily constructive, and a sufficient degree of turn-taking occurs (e.g., asking and answering comprehension questions with a partner, Chi and Wylie, 2014). Not all interactivity is the same. This is even truer for the level of interactivity of computer-based systems (such as intelligent tutoring systems (ITS)). For example, on the question of whether ITS are truly interactive when they require a simple response from the student, Chi and Wylie’s (2014) definition would suggest not, asserting that interaction is student-centered and applies to the kind of responses that a student makes. For example, when a student selects an answer from a menu of choices, selecting is only active in their taxonomy in that the student does not generate a product. This means that responses to a so-called interactive computer system can also be classified as *passive*, *active*, *constructive*, and *interactive*. This differentiation helps to distinguish new emerging SPAT from already existing ITS.

In the context of a constructivist learning paradigm and the ICAP-framework, SPAT should try to imitate the described educator’s behavior working more as a facilitator, coach, and guide for instruction rather than a knowledge mediator. It should try to build complex dialogues with students, bringing them from a *passive* to an *interactive* learning behavior. Its use in education may help to shift the focus from knowledge-as-possession to knowledge-as-construction, and from learning as outside-guided to learning as self-guided. The penetration of SPAT into the learning process has the potential to help students construct their own knowledge and support them within their own ZPD (Graesser and McDaniel 2017).

2.2.1 Implications of Constructivism for Transforming Online and Blended Learning Environments

In online and blended learning environments, educators meet students with very different characteristics such as background knowledge, experience, motivation levels, and interests. According to the constructivist view of learning, students construct new knowledge by creating unique connections with their own experiences, meaning that students and educators both play a role in facilitating and generating knowledge (Jonassen 1991). Moreover, constructivism encourages active rather than passive learning and the use of social learning activities, which can both be mediated through technology. These constructivist views of learning should be integrated in online and blended learning environments to encourage active, collaborative, and responsible students. In online and blended learning environments in particular, where students are not usually in close physical proximity to the educators, there is a stronger need to design technology-supported constructivist learning environments wherein students need individual support and are required to work socially with each other, and to move the teacher from podium to sideline, from leader to coach, from purveyor of knowledge to facilitator of personal meaning making (Romiszowski and Haas 1989). However, educators do not have enough resources to be the individual facilitator or coach for all of their students. Hence, online and blended learning environments should exploit further the potential and capability of IT to foster individual support, interactive communication, and collaboration (Jonassen et al. 1995). Jonassen et al. (1995) asserted that a constructivist approach to knowledge construction and learning can be well supported in online and blended learning environments—intelligent tutoring systems such as SPAT, for example, can offer educators alternative approaches that provide individual support and facilitate learning. These constructivist learning environments and tools can replace the deterministic educator-controlled model of distance instruction.

Constructivist design principles can assist understanding, guiding educators' decision making so that decisions regarding the application of technologies are made more reflectively and rationally. Educators who clarify and articulate their philosophical position about the use of technologies in the learning process know what they are doing and why they are doing it, as they use technologies to encourage constructivist learning (Anderson and Kanuka 1999). In online and blended learning environments, it is much easier for educators to follow an instructional systems design model and feel that it is correct and effective because the literature says so rather than grapple with the complexities of the "ill-structured" world in which people must function (Jonassen et al. 1995). This is why most educators use instructional approaches (e.g., static online videos) to transfer their content. A major problem with this instructivism, then, is that it discounts the reality of an ambiguous, complex, and continually changing world. Consequently, learning activities must be presented in an ill-structured way that reflects the complexity of functioning in a changing world after the course has ended. The educator is a guide, helper, and partner, and students must construct knowledge

by themselves; the content is secondary to the learning process, and the source of knowledge lies primarily in experiences. The educator's role is not to provide information—they must create the conditions within which learning can take place. The focus of education is not the content, it is the process.

Most educators are aware that constructivist learning environments are more effective, but they remain instructive because they do not know how to create these learning environments with the given resources (Bakki et al. 2019). This dissertation shows how SPAT can help to integrate constructivist principles into online and blended learning environments with the given financial and organizational restrictions to address the problem of individual support in online and blended learning environments. To successfully integrate SPAT into learning environments, it is necessary to know the different components of online and blended learning environments and how they are linked together. In Subsection 2.2.2, I explain how constructivist learning environments can be built and what role SPAT can play in this.

2.2.2 Building Constructivist Learning Environments with Technology

The main challenge faced by educators creating TML environments lies in creating learning, monitoring, and tutoring environments to support students so that they do not feel overwhelmed. Designing such environments is challenging, since it requires effective collaboration between educators and students throughout the course. To be effective, learning methods should be designed in accordance with the underlying epistemological perspective (Gupta and Bostrom 2009). These complexities require mechanisms and methods that can guide and support the design of online and blended learning environments (Bakki et al. 2019). The area of learning-scenario building aims to help educators creating learning environments. Learning-scenario building is a process of modeling learning situations and is an engineering activity that lies at the heart of educators' activities (Bakki et al. 2019). In particular, educators need to find and create adequate resources and activities that are suited to the objectives of the course, make them easily reachable, and integrate these resources/activities into a relevant learning environment (Bakki et al. 2019). This process defines the expected course and specifies the roles, activities, organization, and environment resources and tools needed to carry out these activities (learning, tutoring, monitoring, etc.). These mechanisms must be simple, understandable, and easily applicable to a wide range of fields of knowledge.

Bakki et al. (2019) proposed a learning scenario model intended to help educators create online and blended learning environments based on constructivist principles. The model is centered on the students' learning activity and is divided into four different levels: *environment level*, *scenario level*, *activity level*, and *actor level* (see Figure 5). Within the scenario level, the basic element of an online or blended learning course is the "scenario" class. This class is characterized by its

name, a launch date, the organizational owner of the course and the general objective. The scenario level is further structured into two levels: the learning unit and the lesson. The learning unit may contain one or several lessons.

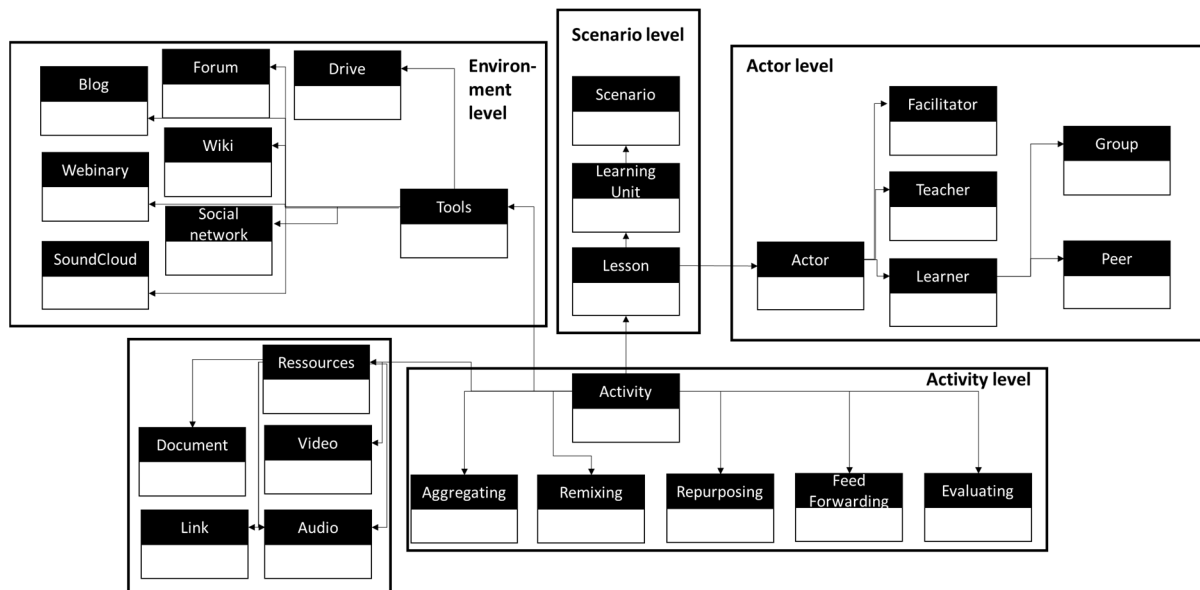


Figure 5. Learning Scenario Model

Source: Bakki et al. (2019)

A lesson is a static entity, defined by the educator at the start of the course, which aims to structure learning and evaluation activities throughout the course. A lesson can be identified by a name, a duration, and a sequencing number. A lesson is composed of learning activities. A learning activity is in the center of the model and should represent the focus of the learning process proposed by the constructivist view of learning. The activity plays the most important role in a learning environment and can be divided into the following five subprocesses:

- *Aggregation*, which involves reading, choosing, filtering and selecting from the content, generating that which is most relevant and appropriate to the student's personal learning objectives. *Remixing*, which is defined as the action of interpreting data, resources and information collected during the aggregation phase, and searching on the web for appropriate resources to address stated problems. This also involves participating in social networks and videoconferences.
- *Repurposing*, which aims to encourage students to create new educational resources. Repurposing activities encourage students and support them in a production process to solve a stated problem, either individually or in groups.
- *Feed Forwarding*, which consists of encouraging students to share their knowledge through the network (e.g., submitting an assignment for a peer review).

- *Evaluating*, which aims to defend the own solution in front of others and receive feedback on the own-knowledge construction process and problem solution.

For learning activities to be operationalized, the environment level is needed, containing resources and tools to support learning activities and support the knowledge construction process. This level describes learning resources and tools as well as their links with activities. This dissertation deals specifically with one type of learning tool, namely SPAT, and how students can best use such tools for their learning activities in online and blended learning environments. The activity level is regulated by an actor, who may either carry out or propose the activity. The actor level defines different roles to be played in learning environments and the types of possible interaction. Actors may have three roles: the student, the educator, and the facilitator. The student defines their personal objective for the learning activity and can interact with peers and groups.

In addition to the epistemological perspective, the learning method structures are a part of the TML framework. Within the TML framework proposed by Gupta and Bostrom (2009), IT is one part of learning method structures and can be defined as a combination of structures that guides individuals to achieve learning outcomes. De Sanctis and Poole (1994) stated that IT can be described in terms of its features and dimensions. Structural features are the specific capabilities offered by an IS, and the dimensions answer the question of what these features do for the actor/students. It is not enough to only understand the IS features—the actors must understand the interactions of all parameters within a learning environment. This is why all levels of a learning environment should be considered. Gupta and Bostrom (1994) noted that it is not the technology or the features of technology that are important, but rather the structural dimensions that the technology provides, which influence learning effectiveness. This is why this dissertation investigates how different DPs of SPAT affect students' learning processes in different learning activities and environments. In Section 2.3, I define my understanding of SPAT and delimit it to other tutoring systems.

2.3 Learning Method Structure: Smart Personal Assistant Technology

Recent technical developments in the area of natural language understanding and NLP creates new use cases for new technologies. SPAT (e.g., Amazon's Alexa, Google's Assistant, or Apple's Siri) can be a promising solution to the problem of offering individual support across different domains. By 2022, it is predicted that SPAT will be running on 870 million devices, which significantly increases the exposure of many to this technology (Botsman 2017). SPAT is a technology that uses inputs such as the user's voice, vision (images), and contextual information to provide assistance by answering questions in natural language, making recommendations, and performing actions (Hauswald et al. 2015).

SPAT is typically used in dialog systems for different reasons, including customer service and information acquisition. (Serban et al. 2017). SPAT has a history that goes back to 1964, when the first chatbot ELIZA was created. ELIZA analyzed input sentences and created its response based on reassembly rules associated with a deconstruction of the input (Weizenbaum 1966). More than 30 years later, the chatbot A.L.I.C.E. was developed using Artificial Intelligence Markup Language, which is based on categories containing a stimulus, or pattern, and a template in the response. Category patterns are then matched to find the most appropriate response to a user input (Wallace 2007). In 2008, Cleverbot was launched and, unlike other dialog-based systems, its responses were not pre-programmed. Instead, it learned directly from human input, where a user would type in a comment or question and Cleverbot would find all the keywords or an exact phrase matching the input (Gehl 2014). Nowadays, SPAT devices are able to capture a wide range of user cases and are able to steer the user in a desired direction.

The remarkable development of SPAT in recent years has resulted in three dimensions for classifying them. Dimension one, “building approaches”, distinguishes between retrieval-based models and generative models. Retrieval-based models use a repository of predefined responses and some kind of algorithm to pick an appropriate response based on the input and/or context. This algorithm can be very simple, such as a rule-based expression match, or more complex, such as machine learning classifiers. Generative models do not reply with predefined responses; instead, they generate responses to the input with the help of machine-learning techniques. Both approaches have advantages and disadvantages. Retrieval-based methods are easier to build and do not make grammatical mistakes, but they are unable to handle unseen cases where no predefined response exists. Generative models are able to refer back to contextual entity information mentioned earlier in the conversation. This causes the SPAT to seem more human and makes it capable of longer dialogs (Liu et al. 2016). The second dimension is related to the input mode of the SPAT. Newly developed SPAT tends to prefer speech over text as input for a conversation (e.g., Amazon’s Alexa or Google’s Assistant). Both methods have advantages. Speech input enables the user to have a more natural and convenient conversation with SPAT compared to text input, which enables the user to trigger conversations in places where sound might not be socially accepted, such as schools and offices (Abu Shawaar and Atwell 2007). The third dimension addresses the inclusion of contextual information. In human-to-human conversations, people keep track of what the dialog partner has said or done in order to use that information for further conversations. For example, humans can ask “how did it go?” if the dialog partner has given an important presentation earlier in the day. Humans automatically know that this question refers to the presentation, whereas SPATs struggle with the context. Hence, SPAT developers try to embed contextual information such as linguistic (e.g., something mentioned earlier in the text) and physical context

(e.g., date/time, location, user information) in order to select the right responses (Britz 2016).

SPAT represents a novel class of IS that is characterized by high degrees of interaction and intelligence. There are various terms describing these assistants—for example chatbot, conversational agent, SPAT, virtual assistant, digital assistant—which are all based on the idea of interacting with users via natural language (Maedche et al. 2016). This dissertation refers to SPAT as an AI-based system embedded in personal technologies designed to assist users by interacting in a text or voice-based conversation (McTear et al. 2016).

In education, experts anticipate AI in education to grow by 43% in the period 2018–2022 (Educause 2019). With heavy investment by private companies such as Google, which acquired European AI start-up Deep Mind for \$400 million, and non-profit public–private partnerships such as the German Research Center for Artificial Intelligence (DFKI), it is very likely that this wave of interest will soon have a significant impact on educational institutions (Popenici and Kerr 2017). Luckin et al. (2016) described three categories of AI software applications in education that are available today: personal tutors, intelligent support for collaborative learning, and intelligent virtual reality. This dissertation focuses on personal tutors and in particular on ITS. ITS can be used to simulate one-to-one personal tutoring. Based on learner models, algorithms, and neural networks, ITS can make decisions about the learning path of an individual student and the content to select, provide cognitive scaffolding, and help to engage the student in a dialog. ITS have enormous potential, especially in large-scale blended and online teaching courses, which run modules involving thousands of students, where human one-to-one tutoring is impossible (Zawacki-Richter et al. 2019). There are already many types of ITS out there that have demonstrated that they can keep up with human tutors (D'Mello et al. 2010). For example, AutoTutor and its derivations have been very effective as a learning technology (Nye et al. 2014). AutoTutor has produced learning gains that are on average around 0.8 standard deviations above human controls who read static instructional materials for an equivalent amount of time (Graesser et al. 2014). Nevertheless, ITS have not yet managed to become widespread in learning environments. It seems that the adoption of these systems by teachers is one of the main barriers (Nye 2014).

Researchers agree that SPAT represents a novel class of ITS that calls for a new approach to studying its influence (Canbek and Mutlu 2016; Graesser et al. 2018c). Compared to other subtypes of ITS, the novelty of SPAT lies in two major areas that potentially fundamentally affect the way they can support students: The underlying technology of SPAT, which allows students to interact with SPAT in a new way; and the easy accessibility and creation of SPAT for students and educators. In the past, ITS typically relied on rigid behavioral patterns, where action-condition rule-based, Bayesian network, and data mining were the most frequent AI techniques applied (Mousavinasab et al. 2018). Most of them were only able to respond to simple requests by matching the user's input against a set of stored patterns (McTear et al. 2016). There were little or no machine learning techniques incorporated in the system. However, based on advances in natural

language understanding, NLP and machine learning techniques such as deep learning, as well as higher computing power stored in clouds, these systems are now better able to detect and classify the intent of users. Current SPAT chiefly uses deep learning algorithms that help to model a large number of domains and transfer that learning to a new domain or skill. These algorithms help to better recognize students' utterances without the need of self-generated training data, which other types of ITS often still require (Kloos et al. 2020). One example is Amazon's Alexa, which supports users to carry out everyday tasks via an advanced voice user interface, ultimately acting as their personal assistant. The natural language understanding and processing happens in a cloud hosted by Amazon, and all students require is an app or standalone device such as Amazon's Alexa Echo Dot. SPAT employs text, speech, graphics, haptics, gestures, and other modes in various combinations for communication trying to help students conduct tasks and thereby assisting educators or imitating them (McTear et al. 2016). The design, implementation, and strategies of SPAT employed in education vary widely, which reflects the diverse nature of the evolution of this technology field (Kerry et al. 2009). Interactions between students and SPAT are usually textually mediated (e.g., Song et al. 2017), with the SPAT showing questions or hints and students clicking buttons or typing responses on the keyboard. Some systems use embodied SPAT (e.g., Perez 2017) capable of displaying emotions and gestures, whereas others use simpler avatars (e.g., Kerly et al. 2008). Voice output, using text-to-speech synthesis, is used in some systems (e.g., Nye et al. 2014), and speech input systems are increasingly viable (e.g., Litman et al. 2006). With recent technological developments in natural language understanding and processing, there is more and more SPAT that engages in largely free interactions with the students resulting in knowledge gains (e.g., Pai et al. 2020). These technical advances enable the systems to offer dynamic scaffolding to students in a new way. SPAT can answer students' questions via voice in real-time letting them assume the role of an individual coach.

The second major aspect that differentiates SPAT from other types of ITS is its accessibility and the provided graphical user interface that empowers educators and students to easily access and create their own SPAT interface. SPAT is increasingly integrated in devices students use every day (such as Google's assistant or Apple's Siri on smartphones, Microsoft's Cortana on desktop PCs, etc.), accessible mostly via one click or voice command, and this is why they have the potential to become daily companions in students' private lives and school career. Furthermore, SPAT providers offer educators intuitive, graphical interfaces that allow them to create their own SPAT without having deep programming knowledge or using authoring tools specially developed for this purpose as with other ITS subtypes. SPAT can be differentiated into two types: built-in SPAT that is included in multi-purpose devices (e.g., Siri for Apple products and Cortana for Windows-based PCs); and standalone SPAT that is included in dedicated devices (e.g., Alexa in Echo, Echo Dot, and Tab dedicated devices). In this dissertation, I implement both types of SPAT by enabling access via different devices.

Given that the design and use of SPAT as a new subtype of ITS in education is an emerging field, most of the studies are explorative and few empirical studies exist that rigorously measure the effect of SPAT on learning performance. Canbek and Mutlu (2016) examined the potential effect of SPAT in learning environments by conceptually investigating potential use cases of different types of SPAT, such as Amazon's Alexa and Apple's Siri. They found that one major benefit of SPAT is that it improves students' listening and speaking skills without the need for human tutors. Dousay and Hall (2018) observed how teachers and administrators perceived the implementation of 90 Amazon's Alexa Echo Dots in four school districts in the US involving approximately 900 students. Teachers and administrators used Alexa to set reminders for such things as activities, events, student dismissals, and student medications. They discovered that one of the biggest challenges of implementing SPAT was the training of teachers and administrators. Jean-Charles (2018) focused more on the perception of teachers regarding using SPAT in their learning environments. In specific, he asked pre-service teachers about their perspective on using Google's Assistant in the classroom. One of the biggest concerns of the teachers was to keep the information secure while allowing the SPAT to gather the data to perform its functions. Moreover, in agreement with Dousay and Hall (2018), Jean-Charles (2018) found that most of the teachers did not feel prepared to use this technology in a meaningful way. Arend (2018) implemented SPAT for a specific learning environment, implementing Apple's Siri during a literacy activity in a third-year college English class. She used a conversation analysis approach and found that one of the big advantages of interacting with Siri was that Siri triggered students to make their thoughts explicit, which helped them to structure their thinking processes. There are also few studies that used systems similar to SPAT to support skill development. For example, Mavrikis et al. (2014) used a speech-based ITS, which helped students to think aloud significantly more than in past interactions and to rephrase their statements to employ mathematical language. Moreover, Damacharla et al. (2019) used Amazon's Alexa for emergency care provider training to help to avoid medical errors, and this resulted in more efficient and effective training.

In sum, SPAT can be seen as a novel subtype of ITS because of its new underlying technology and the facilitated creation and access of these systems. These systems might have the potential to assume the role of an individual coach similar to the role of human educators from a constructivist perspective. As stated in Subsection 2.2.2, it is not only SPA features that are important but the manner in which they interact with other parameters within a learning environment. To better understand how SPAT can influence learning processes and learning outcomes, the dissertation describes the process and outcome perspective of TML in the next two sections.

2.4 Learning Process

To really know how a new type of ITS can influence learning outcomes, it is necessary to take a closer look at how students use them as part of their learning process. The learning process can be viewed as an appropriation of a structuration process in which participants learn and adapt the learning method structures based on their learning goals (Gupta and Bostrom 2009). Scaffolding describes the process of technology appropriation and focuses on how to guide or facilitate students in constructing new knowledge. To better understand how SPAT can help students in their learning processes, it is necessary to understand the mechanisms of scaffolding.

Wood et al.'s (1976) theory of scaffolding emerged as a part of social constructivist theory and was particularly influenced by the work of the Russian psychologist Lev Vygotsky. Vygotsky (1978) argued that we learn best in a social environment, where we construct meaning through interaction with others. His Zone of Proximal Development (ZPD) theory, which states that we can learn more in the presence of a knowledgeable other person, became the basis for the Theory of Scaffolding (Hedegaard 2002). Among other things, this theory states that when learners strive to acquire new knowledge, they need individualized support falling within their individual ZPD. The ZPD represents the potential distance the learner could reach with the help of a more knowledgeable other (Vygotsky 1978). As students advance and become more independent in their thinking, this support can gradually disappear. Four types of scaffold have been identified in the IS and education literature: cognitive, meta-cognitive, procedural, and strategic (Hannafin et al. 2004). Cognitive scaffolds are a mechanism of providing a scaffold using IT to construct new knowledge. Meta-cognitive scaffolds support individual reflection on learning, such as soliciting estimates of current understanding or cuing participants to identify prior related experiences they can reference. Procedural scaffolds help participants make navigation decisions, such as how to utilize available resources and tools. Strategic scaffolds support students by anticipating their interactions, by analyzing, planning, and making tactical decisions, for example. Moreover, Saye and Brush (2002) distinguished between static and dynamic scaffolding. Static scaffolds are static supports that can be anticipated and planned in advance based on typical student difficulties with a task. In contrast to static scaffolds, SPAT tries to use dynamic scaffolds that are situational and requires the system to continuously diagnose the understandings of students and provide timely support based on student responses.

Lepper et al. highlighted three significant aspects of a scaffolding interaction: (a) it supports learners in the achievement of tasks beyond their unassisted capacity; (b) when the support structure is removed, learners continue to function competently on their own; and (c) removing the support structure does not reduce learning or functioning—instead, learners continue to function at the elevated plane reached via scaffolding. One prominent scaffolding strategy is called *directed lines of reasoning*, in which educators ask a series of directive questions to help

the learner dive deeper into the topic (Mills et al. 2004). If the learners' answers are correct, the educator advances to the next question. Otherwise, the educators try to scaffold the learners' understanding with additional questions that meet the learners' individual ZPD for the targeted concepts (Albacete et al. 2018).

This dissertation mainly focuses on metacognitive and cognitive scaffolds and the *directed lines of reasoning* strategy to help students to construct new knowledge similar to the role of an educator within a constructivist learning environment. New emerging SPAT such as Google's Assistant or Amazon's Alexa have the capacity to become a "more knowledgeable other" for students. They can build a dialogue with students by providing dynamic scaffolds and react to students' questions and responses in a new way similar to a human tutor. From an ICAP-framework perspective, SPAT might have the capacity to bring students into an interactive learning mode resulting in a deeper involvement with the learning material.

2.5 Learning Outcomes

Learning outcome is a multi-faceted construct. How to classify and measure learning outcomes resulting from a learning process has been discussed extensively in past research. Cognitive psychologists and neuroscientists provide evidence supporting a distinction between declarative and procedural knowledge (Berge and van Hezewijk 1999). Declarative knowledge refers to factual knowledge such as knowing that "The Second World War started in 1938". In contrast, procedural knowledge refers to process-related knowledge such as knowing how and therefore being able to demonstrate "how" in a statement such as "knowing how to structure one's own learning process". These two types of knowledge are not observable by a third person. Knowledge, which is observable by a third party, can be summarized under psychomotor skills. In addition to these types of knowledge, Goleman (2013) pointed out that there exists a third category named affective knowledge. Affective knowledge means managing one's own emotions and the inner potentials for positive relationships. These inner potentials further motivate us to engage more in some activities than in others. In summary, human learning outcomes can be divided into cognitive, affective, and psychomotor.

Anderson et al. (1999) built on these findings and tried to classify learning outcomes with a matrix consisting of a knowledge and a cognitive process dimension. The knowledge dimension consists of factual, conceptual, procedural, and metacognitive knowledge. In addition to the already described factual and procedural knowledge, they added conceptual and metacognitive knowledge. Conceptual knowledge shows the interrelationships among basic concepts and metacognitive knowledge is the knowledge of cognition in general (see Table 3).

	remem- ber	under- stand	apply	analyze	evalu- ate	create
Factual						
Concep- tual						
Procedural						
Metacog- nitive						

Table 3. Taxonomy of Learning Outcomes

Source: Anderson and Krathwohl (2001)

The second dimension describes cognitive processes. It consists of remembering, understanding, applying, analyzing, evaluating, and creating. This dimension is considered to be a hierarchy in the sense that the six major categories are believed to differ in their complexity, with remembering being less complex than understanding and so on. In this dissertation, I investigate SPAT not only for factual knowledge, but also for the development of conceptual, procedural, and metacognitive knowledge and try to explore whether SPAT can also support the “applying” and “creating” of new knowledge. For example, Section 5.2 describes how I used SPAT to help students apply problem-solving steps, which would address the category of *metacognitive knowledge* and *apply*.

3 Philosophical and Methodological Background

“Science is the attempt to derive knowledge from facts through certain methods in a systematic and organized way” (Recker 2013, p. 11)

In this chapter, I outline the philosophical and methodological background that informed my research strategy. This includes the underlying research epistemology that influenced the chosen research methods and processes. Moreover, this section outlines the methodological basis and details concerning literature reviews and qualitative and quantitative methods, as well as design science research.

3.1 Philosophy of Science and Research Epistemology

The philosophy of science is a subdiscipline of philosophy and is concerned in particular with the possibilities as well as boundaries of scientific knowledge gain (Döring and Bortz 2016). Research epistemology is about how to create scientific theories and distinguish them from false propositions. This chapter is especially relevant to understanding how the dissertation creates and validates theory. In the following sections, I will briefly outline how different kinds of epistemology influenced the used research methods and processes and refer to the core contributions.

3.1.1 Information Systems Research

In the context of IS research, Recker (2013) suggests a distinction between natural sciences and social sciences. Natural sciences are concerned with natural phenomena and are often considered as the more “exact sciences”, whereas social sciences have far less precision than the natural sciences. A theory from the field of the natural sciences would be, for example, the properties of gravity, while a theory from the field of social sciences would be the nature of individuals’ adoption of technology. Döring and Borth (2016) further distinguished the non-empirical and empirical sciences, as well as the engineering sciences. Table 4 displays a classification of scientific disciplines and their subgroups.

Non-Empirical Sciences		Empirical Sciences			
Formal Sci-ences	Humanities	Social Sci-ences	Natural Sci-ences	Engineering Sciences	
• Philosophy • Mathematics • Etc.	• Theology • Legal Science • History • Literary Studies • Linguistics	• Psychology • Medicine • Educational Science • Sociology • Economic Sciences	• Physics • Chemistry • Biology • Earth Science • Astronomy • Etc.	• Mechanical Engineering • Electric Engineering • Civil Engineering	

	• Media Science • Etc.	• Communication Sciences • Etc.		• Process Engineering • Computer Science • Etc.
--	---	--	--	---

Table 4. Classification of Scientific Disciplines

Source: Döring and Bortz (2016)

IS research can be considered as a multi-disciplinary research domain. The majority of IS research can be categorized as social science, since it investigates IS used by individuals, organizations, and other stakeholders (Recker 2013)—for example, understanding how SPAT influences individuals’ learning processes. However, IS research is also linked to formal sciences and the engineering sciences. In recent years, several research papers have been published that could be categorized as engineering sciences. This subgroup is especially important for this dissertation, since the study investigates how SPAT can be designed from a more engineering-inspired perspective. This engineering angle is also highlighted by IS scholars and practitioners such as Benbasat and Zmud (2003, p. 186), who wrote:

“IS scholars and IS practitioners strive to increase their collective understandings of (1) how IT artifacts are conceived, constructed, and implemented, (2) how IT artifacts are used, supported, and evolved, and (3) how IT artifacts impact (and are impacted by) the contexts in which they are embedded.”

Benbasat and Zmud (2003) promoted the notion of IS research as both a social and engineering discipline, which contradicts a simplified understanding of IS as a social science. Based on this understanding of IS research, it is not yet clear which epistemological beliefs influence the used research methods and the corresponding research contributions. Hence, in the following I outline central epistemological streams that form the basis of this dissertation.

3.1.2 Theoretical Reasoning and Major Streams in Information Systems Research Epistemology

Research epistemology deals with the question on how to create theories based on certain methods that form a new contribution to the greater body of knowledge. When focusing on the reasoning between data and theory, inductive and deductive reasoning can typically be differentiated, while approaches such as abduction or theorizing through design and action also exist (Döring and Bortz 2016). In inductive reasoning, the researcher relies on specific data to derive patterns that are used to theorize. This means that an inductive research process builds generalizable theories based on specific observations from data. This kind of reasoning can be found in qualitative research approaches (Urquhart et al. 2010).

In contrast to this, deductive reasoning consists of reasoning from theories that explain empirical data. In such a research process, hypotheses are logically derived from theoretical assumptions, which are then proven with empirical data. Deductive reasoning helps in the preliminary confirmation or criticism of a theory proposed to explain a hypothesized relationship. This kind of reasoning can be found in quantitative research approaches. Abductive reasoning starts with an observation and then seeks to draw the simplest and most likely conclusion from the observations. In contrast to inductive reasoning that looks for patterns in data, abductive reasoning theorizes on the basis of inapprehensible values in data. These values form the basis for creating explanative hypotheses and theoretical advances, especially in the qualitative paradigm (Döring and Bortz 2016). Finally, theorizing through design and action is a more distinct approach for theoretical reasoning with the IS discipline (Mueller and Urbach 2017). With the help of interventions in the “real world”, researchers try to change relationships between variables that are subject to the theorization process. If an intervention has an impact on the theoretically desired outcome variables, theorizing through design and action helps to gain a better understanding of the underlying phenomenon of interest.

Döring and Bortz (2016) distinguished three major paradigms of research epistemology that use the above-described logical reasoning processes: (a) *critical rationalism and the quantitative paradigm*; (b) *the qualitative paradigm and social constructionism*; and (c) *the mixed-methods paradigm and pragmatism*. This dissertation and the corresponding research questions make use of all of these paradigms with a focus on *the mixed-methods paradigm and pragmatism*. In the following subsection, the three paradigms of research epistemology are briefly explained.

3.1.2.1 Critical Rationalism and the Quantitative Paradigm

Critical rationalism is the philosophy first developed by Anglo-Austrian philosopher Karl Popper (1902-1994) in his work “The Logic of Scientific Discovery” (Popper 1959). Popper took “falsification” as the distinguishing characteristic of science: universal theories are never verified or confirmed with any positive probability, they can only be falsified (Ormerod 2009). His work is an alternative draft to the predominant empirical positivism that assumes that through the collection of empirical data and inductive reasoning, generalizable theories can be developed. In critical rationalism, theories are first formulated by the researcher. Second, researchers formulate a hypothesis that should confirm the theory using empirical data (Popper 1959). Finally, a preliminary confirmation or falsification of the hypothesis and associated theory is established (Weber 2012). When a theory and its hypothesis is confirmed many times through replication studies, it can be considered as “proven” (Weber 2012). For example, the widely acknowledged technology acceptance model (TAM) as a very robust and generalizable native IS theory can be considered as a proven theory (Venkatesh and Bala 2008).

However, this theory can be rejected or falsified in some cases and thus further developed.

Critical Rationalism argues that a falsification of the theory can arise from two reasons: From the theory itself but also from the data derived from the respective measurement instrument. An example of this would be a poorly designed experiment for gathering insights regarding the adoption of SPAT in learning environments. If the experiment has too many confounding variables (no randomization, no manipulation check, etc.), the results of the data analysis may lead to rejecting the underlying hypothesis of the model as well as the theoretical arguments. From a rationalism perspective, the falsification of theory does not prove that the theory itself is weak, and this is why this epistemological stream is called critical realism, because of the critical question (principle of “criticism”) of assumptions. A purely deductive critical rationalist stance on the basis of a falsification approach might not be fully suitable for moving IS research forward (Merikivi et al. 2018). Consequently, other approaches such as induction are needed, and their importance should not be underestimated.

3.1.2.2 Social Constructionism and the Qualitative Paradigm

In contrast to critical rationalism and the quantitative paradigm, social constructionism and the qualitative paradigm differ widely in terms of research designs, sampling, data collection and analysis methods, and a different understanding of the perceived role of the researcher (Creswell 2014; Döring and Bortz 2016). The epistemological basis of the qualitative paradigm and social constructionism is that research on social reality is completed using subjective interpretations (Orlikowski and Baroudi 1991). The main goal of interpretations is "to understand how the members of a social group realize and endow their particular realities with meaning through their participation in social processes, and to show how these meanings, lifestyles and intentions of the members contribute to constituting their social action" (Orlikowski and Baroudi 1991, p. 13). While positivist researchers think of uncovering an objective reality through inductive thinking, interpretation research assumes that gaining knowledge requires an understanding of "how practices and meanings are influenced by language and tacit norms shared by people working towards a common goal" (Orlikowski and Baroudi 1991). A key aspect of the qualitative paradigm is the circularity and flexibility of the research process, also called the hermeneutic circle, which suggests "that all human understanding is achieved through iteration, taking into account the interdependent meaning of the parts and the whole that they form" (Klein and Myers 1999, p. 72). The interaction between the researcher and the subjects is a crucial point that must be taken into account when considering the data collected (Klein and Myers 1999). For example, this dissertation details qualitative interviews in Section 5.2 to gain more insight into how students perceive interactions with SPAT.

3.1.2.3 Pragmatism and the Mixed-Method Paradigm

In social sciences, it is practical to combine quantitative and qualitative data and the according research methods in a method mix (Venkatesh et al. 2013; Venkatesh et al. 2016). However, there are different understandings of how quantitative and qualitative methods should be integrated into a holistic research process (Döring and Bortz 2016). This research paradigm is neither inductive nor deductive. A mixed-method research process is based on pragmatism, a concept originally developed by Charles Sanders Pierce (1839-1914), John Dely (1859-1952), and William James (1842-1910) (Döring and Bortz 2016). Pragmatism states that researchers should use whatever philosophical and/or methodological approach is best for a particular research project (Yvonne Feilzer 2010).

Pragmatism has shaped IS research to a great extent, and it opens the doors for a mixed-method approach and also for design science research. The pragmatic paradigm argues that it is not of primary importance to generate knowledge that is true in some sense or reconstructs the perspective of the study objects (Yvonne Feilzer 2010). Rather, the pragmatic perspective suggests that the theories developed should meet the criterion of usefulness in solving practical problems (Döring and Bortz 2016). In line with this view, this dissertation develops theories that first explain how SPAT can be developed and implemented in TML and then demonstrates how SPAT can solve problems and challenges in online and blended learning environments. This means that knowledge contributions are not limited to understanding, explaining, and predicting key phenomena, but that they also emphasize constructive knowledge contributions that are prescriptive, normative, and prospective (Goldkuhl 2012).

3.1.2.4 Implications

Drawing on this discussion about epistemological paradigms, my dissertation mainly uses design science research methods to solve formulated problems and challenges of SPAT in TML research. Within the design science research process, this dissertation also uses quantitative, qualitative, and mixed research methods. For example, I adhere to the quantitative paradigm and its epistemological grounding when I experimentally evaluate SPAT in different learning environments. When investigating how SPAT effects the student learning process, I use qualitative methods to gather a deeper understanding. The development and evaluation of design principles and the SPAT platform can be assigned to the pragmatic worldview. To summarize, this dissertation cannot be assigned to only one paradigm. Instead it uses a diverse worldview to better answer the proposed RQs that will in consequence lead to a more conclusive and integrated solution to the proposed problems and challenges in SPAT research. Now that the epistemological orientation of the dissertation has been clarified, I will now explain the research methods used in the upcoming sections.

3.2 Literature Review

“They serve as foundations and frameworks for research projects because they help to develop an understanding of a domain and explain the topic under study.” (Vom Brocke et al. 2015, p. 206)

In line with this quote, literature reviews are a necessary first step in structuring a knowledge domain and recognizing the state-of-the-art, especially when it is as diverse as the IS field and TML in particular (Benbasat and Zmud 2003). TML is situated at the intersection of education, psychology, and IS research. This leads to a great variety of relevant publication possibilities in different outlets and databases, which makes it difficult to keep track (Vom Brocke et al. 2015). In addition, each community has its own notions for SPAT, which makes it even more difficult to distinguish SPAT from other technologies and to keep track of the current developments.

A literature review is a systematic, explicit, and reproducible method for the identification, assessment, and synthesis of existing literature that represents the wider body of knowledge (Vom Brocke et al. 2015). A literature review is not only about summarizing existing research, but also about pinpointing gaps and developing a research agenda as well as theoretical propositions (Webster and Watson 2002). Typically, a literature review involves searching for relevant material on literature databases that enable researchers to access scholarly research by using keywords that relate to the phenomena of interest (Webster and Watson 2002). The choice of keywords and the chosen databases have a high impact on the outcomes of the review. According to Webster and Watson (2002), a literature review should focus on concepts rather than authors, since searching for authors falls short when seeking to identify new research streams or develop a new theory that is related largely to concepts and constructs (Webster and Watson 2002).

To answer RQ1, I conducted a systematic literature review concerning the state of SPAT in TML research resulting in future research topics that need more attention. The review was based on Webster and Watson (2002) and vom Brocke (2015). According to the TML framework proposed by Gupta and Bostrom (2009), I analyzed the current state of the art of SPAT from an input, process, and output perspective. This multi-perspective helps to balance the overly technocentric view by including the process and outcome view. The literature review followed the five stage process suggested by Arksey and O'Malley (2005) consisting of a) identifying initial research questions; b) identifying relevant studies; c) making the study selection; d) charting the data; and e) collating, summarizing, and reporting the results. More details about the approach and concluding findings can be found in Section 4.1 of this literature review. The literature review provides the base for my further research plan. In the following, I describe the methodological basis for the quantitative methods used within my all-embracing design science research approach.

3.3 Quantitative Methods

“The true method of knowledge is experiment.” (Blake 2007, p. 56)

Quantitative methods based on the quantitative research paradigm include a huge variety of different approaches. This dissertation focuses on experimental research studies as well as a CA study. Each research method has its pros and cons, and identifying the right method is linked closely to the research question (Leimeister et al. 2005). To evaluate and choose the right method, McGrath (1995) proposes three maxims, which cannot all be achieved at the same time:

- *Generalizability*: Generalizability with respect to the underlying population is related to the extent to which the results have to be generalizable to the related population.
- *Precision*: Precision relates to the way in which variables related to outcome variables can be controlled and measured, and the extent to which variables can be manipulated and disturbing variables switched off.
- *Existential realism*: Existential realism relates to the participants, the context in which their behavior is observed, and the extent to which the participants' behavior meets realistic conditions.

There is no single method that is able to maximize on all three dimensions—researchers can either make the most of one of the dimensions and reduce the other two or achieve an uneasy balance between two of the dimensions and fail to a higher degree on the third dimension (Dennis and Valacich 2001). This study sought to choose research methods that together were able to maximize on all dimensions by using laboratory (see Section 5.3) and field experiments (see Section 5.2). Laboratory experiments maximize precision by controlling certain variables most of the time and switching off disturbing variables—for example, by controlling the usage time of SPAT. However, because of limited samples they are not fully generalizable and, when conducted in an experimental setting, fail to meet realism requirements for the participants. The latter issue was addressed in my research by field studies in which I sought to implement SPAT in real-world learning environments. In the following subsection, I describe experimental research methods and the adopted Conversational Analysis (CA) approach.

3.3.1 Experimental Study

Experimental studies are used in IS research predominantly whenever it is the goal to provide empirical evidence for a theory while meeting the goals of precision and control. Studies that use an experimental research design investigate cause and effect relationships and, at their most basic, usually consist of an experiment group that receives a treatment and a control group that does not re-

ceive the treatment. By this means, the experimenter can manipulate the independent variables and can control confounding variables that could affect the experiment.

There are several choices that have to be made when creating an experimental design. First, experimental design choices can be differentiated according to different criteria. Researchers have to make decisions related to a) the random assignment of participants; b) the measurement of covariates; c) the treatment administration; and d) the observation of pre- and post-test measures (Kirk 1982; Recker 2013).

In its simplest form, true experimental designs involve a random group assignment: one treatment and one control group suitable for testing the causal effects of a single independent variable that is manipulated with an according treatment (Creswell 2013). I implemented this design in the laboratory experiment described in Section 5.3. Randomization in true experiments mainly serves three purposes: it serves the distribution of individual characteristics of study participants over the treatment manipulations and control group in a way that these characteristics do not bias the experiment outcomes; it permits the computation of unbiased error effect estimates; and it guarantees that error effects in an experiment are statistically independent. However, in the TML domain, full randomization is not always possible—for example, in experimental studies that are carried out in the field with real-world cases such as school classes (Kirk 1982). These limitations applied to the study described in this dissertation. Experiments that consider the key characteristic of experiments but not the full randomization of study participants are called quasi-experiments (Gribbons and Herman 1996). For example, the conducted field experiments described in Section 5.2 are quasi-field experiments and, moreover, pre-tests or covariate measures can be incorporated. Pre-tests help to identify whether the treatment and control groups have similar preconditions and also allow the calculation of learning gains. I used pre-tests in the field experiments in Section 5.2 and the laboratory experiment in Section 5.1. Covariates are continuous variables that are expected to change with the outcome variable of a study. For example, I investigated learner engagement as a covariate when testing knowledge gains as described in Section 5.3. In contrast to this, factorial designs involve the manipulation of two or more factors, and each sub-division of a factor is called a level. This design type enables the examination of main effects of the independent variables as well as the interaction effects on the dependent variables (DV). For example, I investigated different SPAT design principles within one experiment, as described in section 5.3 (Bhattacharjee 2012). Factorial designs are notated with numbers. For example, the above described experiment would be considered as a 2 x 2 factorial design. These more complex experimental designs also involve more experiment groups—four groups in a 2 x 2 experiment, for example. These designs require larger sample sizes for each group to ensure statistical power, and these should be calculated a priori (Cohen 1988; Hair et al. 2014). Furthermore, experiment design types

also vary in terms of their treatment administration. When participants receive more than one treatment, it is called a repeated measure design (Kirk 1982). A popular repeated-measure design is the crossover study. Crossover designs are common for experiments in many scientific disciplines such as psychology and education (Kirk 1982). This study uses such a design in the laboratory experiment described in Section 5.3 in which participants were able to interact with two different types of SPAT.

3.3.2 Quantitative Data Analysis Methods

3.3.2.1 Assessing Differences across Groups

In TML experimental research, researchers often use analysis methods for assessing the differences between groups to draw inferences. In this dissertation, these analysis methods are mainly used to assess the impact of specific SPAT design principles such as scaffolding mechanisms and to compare the use of SPAT with traditional learning methods. For this purpose, I used t-tests, AN(C)OVAs (ANalysis Of (Co-)Variance) and MAN(C)OVAs (Multivariate ANalysis Of (Co-)Variance). The right choice of statistical test depends on a variety of criteria such as the experimental design and the dependent, independent, and control variables as well as the data coding and the distribution of variable scores (Recker 2013).

The t-test helps to examine the differences in mean values between two groups. The t-test can be used, for example, to determine whether the means of two sets of data are significantly different from each other (Mertens et al. 2017). Statistics use one-tailed t-tests for directional hypothesis where researchers know a priori if a mean should be higher than the other (Bhattacharjee 2012). For a non-directional hypothesis, a two-tailed t-test is used because one only tests the difference and not the direction of the difference (Bhattacharjee 2012). There are three different forms of t-tests: the independent-samples t-test (examination of two independent groups that are not influenced by each other); the paired-samples t-test (the investigation of two dependent groups at different times); and the one-sample t-test (comparison of two groups from one sample). To implement a t-test, the following steps have to be performed: (1) comparing the sample means and SDs of the two groups to identify the t-statistics; (2) determining the confidence interval (typically 95% or higher) in the t-table; and (3) checking the probabilities (p-values) of the t-test (Recker 2013).

Researchers use the ANOVA to examine differences between two groups or more (Backhaus et al. 2016). This inferential analysis is characterized by the usage of one DV and one or more independent variables that also give ANOVAs their name, for example a one-way ANOVA for one independent variable (Backhaus et al. 2016). The factorial ANOVA is used for one DV and two independent variables such as a 2 x 2 factorial design with four groups (Mertens et al. 2017). An ANCOVA is a blend of an ANOVA and a regression. It is similar to factorial ANOVA in that it can reveal the additional information that can be obtained by

considering one independent variable (factor) at a time, without the influence of the others (Senn 1994). This is often used in education and especially in pretest-posttest designs to control for the pretest results (Bonate 2000). The analysis of AN(C)OVA with multiple outcome variables is called MAN(C)OVA. More than one DV and one or more independent variables are the main criteria for this analysis method (Backhaus et al. 2016). The computation and process of MAN(C)OVAs is similar to that required for AN(C)OVAS. Table 5 presents a brief overview of the different group comparison methods.

Groups	One DV	More than one DV
One group	One-sample t-test	Multiple one-sample t-tests
Two independent groups	Independent-samples t-test	MAN(C)OVA
Two dependent groups	Paired-samples t-test	Repeated-measures MAN(C)OVA
More than two independent groups	AN(C)OVA	MAN(C)OVA
More than two dependent groups	Repeated-measures AN(C)OVA	Repeated-measures MAN(C)OVA

Table 5. Overview of Group Comparison Methods

Source: Mertens and Pugliese (2017)

In this study, I mainly used t-tests to identify differences between the control and experiment group and to check whether both groups were similar—for example, in terms of gender, age, or personal innovativeness (see Section 5.4). When using a pretest and posttest design, I applied an ANCOVA for two groups and one DV (see Section 5.2) and a MANCOVA for two groups and more than one DV (see Section 5.3).

In addition to the quantitative research methods, this study also uses many qualitative methods that are based on the epistemological principles described in Subsection 3.1.2.

3.4 Qualitative Methods

“I want to understand the world from your point of view. I want to know what you know in the way you know it. I want to understand the meaning of your experience, to walk in your shoes, to feel things as you feel them, to explain things as you explain them. Will you become my teacher and help me understand?”
(Spradley 2016, p. 34)

Whereas the focus of quantitative methods is on the explanation of cause-and-effect relationships, qualitative methods often use inductive reasoning and are responsible for a deeper understanding of complex social issues and the behavior of human beings (Döring and Bortz 2016). As described in the quotation above, qualitative research is a situated activity that locates the observer in the world. Qualitative research methods involve a set of interpretive, material practices that make the world visible. They try to represent the world with data including field notes, interviews, conversations, photographs, recordings, and memos to the self. At this level, qualitative research involves an interpretive, naturalistic approach to the world. This means that qualitative researchers study things mostly in their natural settings, attempting to make sense of, or to interpret, phenomena in terms of the meanings people bring to them (Döring and Bortz 2016). This qualitative research method is particularly useful where constructs of interest are not known in advance and might be unexpected (Döring and Bortz 2016). The two main steps of qualitative research methods are data collection and data analysis (Recker 2013). The data collection includes different techniques such as interviews, observations, documentation, and audiovisual material (Creswell 2014). The interviews can be performed face-to-face or one-to-many (focus groups), and observations may happen directly (passive researcher) or indirectly (researcher as an active participant) (Recker 2013). Documentation relates to public (e.g., newspapers) or private (letters and audiovisual material from websites, social media, etc.) documents (Creswell 2014). For this dissertation, I collected qualitative data through interviews and focus groups to gather requirements of SPAT design and to formatively evaluate SPAT, and further qualitative data collected in quantitative surveys.

After collecting the data, the next step is to analyze it with methods such as coding (development of order categories for texts); memoing (recording of memos as a commentary or reflection); critical incidents (identification and examination of a series of events/states); content analysis (semantic analysis of text documents); and discourse and conversation analysis (investigation of the structure of a communication) (Döring and Bortz 2016). While all of these techniques vary in terms of their procedures, goals, and outcomes, this study employed a coding technique to induce themes proposed by Ryan and Bernard (2003). This technique was used to analyze focus group discussions, as detailed in Section 5.2, and to analyze data from a quantitative survey, as presented in Section 5.4. Compared to content analysis methods where keywords are acknowledged only as part of a simple word count, this analysis method takes the context of immediately adjacent words into account. It includes various forms of quantification, word searches, deviant case analyses, and other analytic tools. This approach is well suited to tackling the challenges of working with focus group data, comparing subgroups, and working within a mixed methods project, as described in Section 5.3 (Ryan and Bernard 2003). Specifically, I focus on the keywords-in-context technique in which researchers identify key words or phrases and then systematically search the literature to find all instances of each key word or phrase. Each

time they find an instance, they make a copy of it and its immediate context. Themes are identified by physically sorting examples with a similar meaning into categories (Ryan and Bernard 2003).

3.4.1 User Interviews

In this dissertation, the user interviews employ a User Experience (UX) research method in which the researcher asks participants questions about a topic of interest (e.g., use of a system, behaviors, and habits) with the goal of learning about the topic (Hartson and Pyla 2012). User interviews in this context tend to be a quick and easy way to collect user data, so they are often used, especially in lean and agile environments. They are closely related to the somewhat narrower and more formal human–computer interaction (HCI) method called critical incident technique, which was introduced by John Flanagan (1954). User interviews give insights into what users think about a design artifact (e.g., a site, an application, a product, a process) and can be performed in a variety of settings:

- Before the development of a design, to inform personas, journey maps, feature ideas, workflow ideas
- As part of a contextual inquiry study, enriching it by supplementing observation with descriptions of tools, processes, bottlenecks, and how users perceive them
- At the end of a usability test, to collect verbal responses related to observed behaviors

This study uses user interviews mainly to gather requirements and consequently to inform DPs and DFs. In this context, one important delivery of user interviews is user stories (US) (Cohn 2004). USs are part of an agile approach that helps shift the focus from writing about requirements to talking about the user. All agile USs include a written sentence or two and, more importantly, a series of conversations about the desired functionality. They typically follow a simple template such as this: As a <type of user>, I want <some goal> so that <some reason>. For example, one US I derived in the process described in Section 4.2 was as follows: As a non-tech-savvy educator, I want *to first analyze for which learning goals, contents, and methods I can use SPAT* so that *I can create an effective SPAT by myself*. This US alone is incomplete unless details are added. Details behind the story emerge during conversation with the user, and acceptance tests confirm that the story was coded correctly. In the above example, details that should be added and asked in the interviews include “How do you analyze the learning goals, contents, and methods for a given learning content?” and “How do you decide if and which technology fits your learning environment?” Tests are often written on the back of a US card. One example for a possible test for the above scenario is “Type in a learning goal and then check whether you get a recommendation on how to use SPAT”.

3.4.2 Focus Groups

A focus group is a gathering of deliberately selected people who participate in a facilitated discussion intended to elicit user perceptions about a particular topic of area of interest. Unlike interviews, which usually occur with an individual, a focus group allows and encourages members to interact and influence each other during the discussion and consideration of ideas. Focus groups are distinguishable from one-on-one individual interviews in that participants of focus groups have the opportunity to clarify and modify their ideas through discussion and by challenging other participants (Belanger 2012). These dynamics create newly developed ideas and provide a great opportunity for the researcher to view the participants' commitment to their views (Belanger 2012). The publicly expressed statement of disagreement, conflict, and conviction allows the researcher to interpret a particular participant's meaning. This conviction is not always as clearly expressed in a one-on-one interview setting.

This dissertation used focus group interviews in two different ways. As described in Section 5.2, it used focus group interviews to uncover how SPAT influences students' learning processes. Within the discussion, participants were able to elaborate on their thoughts and create counterarguments that helped to provide a richer picture of how SPAT influences their learning processes. As described in Section 4.2, I also used focus groups to evaluate the translation of requirements to design principles (DPs).

3.4.3 Conversation Analysis

Conversation Analysis (CA) is a social-scientific research method related to the investigation of interaction. CA is an inductive, data-driven method for the discovery and description of interactional practices and organizations of practice observed in naturally occurring social interactions. By this definition, CA can be classified under the qualitative paradigm. CA provides a methodology, a set of analytic constructs, and a collection of established findings about interaction that illuminate the investigation and design of human-computer interactions in a way that transcends technological issues (Norman and Thomas 1991). As one of a number of disciplines that focus on the nature of human interaction, CA is concerned with the description of interactions conducted through talk. CA examines one particular aspect of the process of sense-making in social interaction—the production and recognition of systematic interactional structures (Norman and Thomas 1991).

CA methods seek to preserve the “integrity of the phenomena” using “real-life” tape-recorded interactions. This is then transcribed using a modified orthography that seeks to record detailed features of the conversation. This attention to the detail is motivated by the assertion that it is impossible to decide a priori the important phenomena that might come to light. In this way, features uncovered by detailed transcription are made available for investigation in their own right (Hepburn and Bolden 2013). In this way, phenomena such as the exact points where

speakers start talking (Sacks et al. 1978), the duration and placement of pauses (Sacks et al. 1978), and the nature of gaze and posture (Goodwin 1980)—things that might have gone unnoticed—are investigated. If systems are designed in accordance with the interactional predispositions, expectancies, skills, and abilities of users, as revealed by CA, they will be superior interactive artefacts. There are three powerful reasons that support this view.

First, the formats in which certain actions in interaction appear are natural to us. CA asserts that such knowledge resides at a level below our conscious awareness. If interactive systems are designed in a way that ignores pervasive expectations, users might be forced into conscious speculation, since they would be engaged in interactions that seemed “odd” or “unnatural”. Second, systems designed by principles revealed by CA would be easier and more intuitive to learn, since there would be little in the way of cognitive overheads in finding conscious routes to achieve specific interactional goals. This means that they would be simpler to operate, since interactions would bear some comparison with conversational interactions and, ultimately, would lead to more effective work practices. Third, guidance derived from conversation analysis would provide a more secure and consistent basis for interaction and could be adopted for each new generation of technology. As a principled approach to design, it offers the prospect of portability in a way that enables the user to lead the technology rather than vice versa (Hepburn and Bolden 2013).

As computers become more interactional, conversation analysis as a method becomes more popular in HCI research (Porcheron et al. 2017). New emerging SPAT in particular can be considered as highly interactive and intelligent systems (Maedche et al. 2016). Hoey and Kendrick (2017) proposed a conversation analysis approach that is divided into seven sub-steps and is based on the recorded data: (1) identify a candidate phenomenon; (2) build a collection of cases; (3) analyze each case in the collection; (4) analyze variation in the collection; (5) define the boundaries of the phenomenon; (6) analyze deviant cases and look for normative evidence; (7) produce a formal account of the phenomenon. Each step is further explained in Table 6.

Step	Description
(1) Identify a candidate phenomenon	Most analyses begin with an observation of something in the recorded data. Anything that participants treat as relevant for their interaction may be considered a candidate phenomenon for investigation.
(2) Build a collection of cases	Based on a preliminary description of the phenomenon in hand, the next step is to examine additional audio and video recordings of social interaction to build a collection of cases that will form the empirical foundation of the analysis.

(3) Analyze each case in the collection	After building a collection, the next step is to analyze each case individually. Start by considering the basic nuts and bolts of any interaction: activity, participation, position, composition, and action. Activity is what participants are doing together through interaction. Participation refers to the roles that participants occupy over the course of a given activity. Position refers to where something occurs in the course of interaction. Composition refers to the verbal, vocal, bodily, or material resources that form an action. Action refers to what the talk or other conduct accomplishes in the interaction.
(4) Analyze variation in the collection	The next step is to investigate the variation produced by the phenomenon. The task is to track forms of variation across the collection and sort cases into ad hoc categories such that you can easily compare variants.
(5) Define the boundaries	Standing in opposition to clear cases are boundary cases. These resemble the phenomenon under investigation but can be shown analytically not to be genuine instances of it. The process whereby one identifies such cases and develops criteria to exclude them from the core collection defines the boundaries of the phenomenon (see Schlegoff (1997)).
(6) Analyze deviant cases and look for normative evidence	Deviant cases are a particularly powerful form of evidence for demonstrating the normative organization of a phenomenon. Deviant cases feature a departure from an expected pattern and an observable orientation to it as a departure from the norm (Maynard and Clayman 2003).
(7) Produce a formal account of the phenomenon	The final step in the research process is to produce a formal account of the phenomenon. The criteria that you have developed to identify the phenomenon and its boundaries are an essential part of this account, as is the analysis of the variation in the collection.

Table 6. Conversation Analysis Approach

Source: Hoey and Kendrick (2017)

This study used a conversation analysis approach to uncover hidden interaction patterns of groups interacting with SPAT compared to human tutors, as described in Section 5.1. Using this method, I was able to ascertain how interactions differ and why the identified effects might have occurred.

3.5 NeuroIS Methods

Psychophysiological tools (e.g., eye tracking, skin conductance) and brain imaging tools (e.g., fMRI, EEG) have recently received increasing attention in the social sciences due to their ability to complement existing sources of data with data captured directly from the human body (Vom Brocke and Liang 2014). Neurophysiological tools enable the measurement of human responses when people engage in various activities, such as decision making, or react to various stimuli, such as IT interfaces. The term NeuroIS has been coined to describe the “idea of applying cognitive neuroscience theories, methods, and tools to inform IS research (Dimoka et al. 2012). The promise of NeuroIS is to complement existing research tools with neurophysiological tools that can provide reliable data that are difficult or impossible to obtain using traditional tools, such as self-reported or archival data. The primary advantage of physiological and brain data is that subjects cannot consciously manipulate their responses, since these are not readily subject to manipulation. As neurophysiological data are generally not susceptible to subjectivity bias, social desirability bias, and demand effects, they could complement existing sources of data, triangulate across measurement data, and reduce common method bias by not relying on any single measurement method (e.g., Dimoka et al. 2012). Neurophysiological tools are particularly valuable for measuring IS constructs that people are either unable, uncomfortable, or unwilling to truthfully self-report. This may include sensitive issues (e.g., gender, race, culture, religion); personal issues (e.g., goals or fears); deep or hidden emotions (e.g., habit and automaticity); complex cognitive processes (e.g., cognitive overload); social dynamics (social cognition); antecedents of human behaviors (e.g., beliefs, attitudes, and intentions); and moral issues (e.g., ethics and moral judgments).

Neurophysiological data have the advantage of including continuous real-time measurement that allows the continuous collection of data on a real-time basis while a subject is executing a task or responding to a specific stimulus (something not easily afforded by self-reports or observation). This enables a level of temporal precision that allows a researcher to temporally map the task or stimulus to the neurophysiological response virtually in real-time. Table 7 presents an overview of commonly used neurophysiological tools and the focus of measurement.

Neurophysiological Tools	Focus of Measurement
Psychophysiological Tools	
Eye Tracking	Eye pupil location (“gaze”) and movement
Skin Conductance Response (SCR)	Sweat in eccrine glands of the palms or feet
Facial Electromyography (fEMG)	Electrical impulses caused my muscle fibers

Electrocardiogram (EKG)	Electrical activity of the heart measured via the skin
Brain Imaging Tool	
Functional Magnetic Resonance Imaging (fMRI)	Neural activity measured by changes in blood flow
Positron Emission Tomography (PET)	Metabolic activity measured by radioactive isotopes
Electroencephalography (EEG)	Electrical brain activity measured via the scalp
Magnetoencephalography (MEG)	Changes made to magnetic fields by brain activity

Table 7. Neurophysiological Tools and Focus of Measurement of Neurophysiological Tools

Source: Dimoka et al. (2012)

This study uses Psychophysiological Tools (Skin Conductance Response (SCR), Facial Electromyography (fEMG), and Electrocardiogram) to identify how users react to SPAT DPs. In particular, I measured student's engagement while interacting with SPAT. These data are complemented with posttest measures to uncover possible explanations of identified effects.

3.6 Design Science Research

“The result of the DSR project can be a product or a process; it can be a technology, a tool, a methodology, a technique, a procedure, a combination of any of these, or any other means for achieving some purpose (Venable and Baskerville 2012, p. 142)”

The overall goal of the study described in this dissertation is to design and evaluate SPAT DPs and a SPAT platform based on theoretical insights to further inform the design and use of SPAT in online and blended learning environments. Design science “creates and evaluates IT artifacts intended to solve identified organizational problems” (Hevner et al. 2004) and thus receives ever more attention in the IS community (Gregor and Hevner 2013). In the following subsections, I first introduce the role of theory in design science and then go further into the main design science research processes that were used when developing the SPAT solutions in this study.

3.6.1 Theory in Design Science Research

The fundamental principle of design science research is that the development of artifacts should itself contribute to knowledge and understanding of design problems and their solutions (Recker 2013). Artifacts are central for design science and refer to things as artificially made objects or processes (Gregor and Hevner

2013). As such, artifacts are transformed into material existence—for example, abstract algorithms can be converted into software (Gregor and Hevner 2013). Research has examined design knowledge in terms of design theories (Gregor and Jones 2007) and technological rules (van Aken 2004). Moreover, Gregor and Hevner (2013) came up with the idea of a continuum reaching from the knowledge represented in an instantiation of an artifact, through nascent design theory in the forms of DPs, schemas, and methods, to full design theory. Table 8 presents a framework that categorizes these knowledge contributions. The contribution types above the second level represent a type V theory of design and action when relating to other theoretical contributions in IS research (Gregor 2006).

	Contribution Types	Example Artifacts
More abstract, complete, and mature knowledge ↑ ↑ ↑ ↑ More specific, and less mature knowledge	Level 3. Well-developed design theory about embedded phenomena	Design theories (mid-range and grand theories)
	Level 2. Nascent design theory knowledge as operational principles/architecture	Constructs, methods, models, DPs, technological rules
	Level 1. Situated implementation of artifact	Instantiations (software products or implemented processes)

Table 8. Design Science Research Contribution Types

Source: Gregor and Hevner (2013)

Gregor and Jones (2013) proposed a framework that supports researchers in describing design theories with six core components and two additional components. Table 9 provides an overview of the design theory components.

Component	Description
Core Components	
1) Purpose and scope (the cause finalis)	“What the system is for,” the set of meta-requirements of goals that specifies the type of artifact to which the theory applies and in conjunction also defines the scope, or boundaries, of the theory.
2) Constructs (the causa materialis)	Representations of the entities of interest in the theory.

3) Principle of form and function (the cause formalis)	The abstract “blueprint” or architecture that describes an IS artifact, either product or method/intervention.
4) Artifact mutability	The changes in state of the artifact anticipated in the theory, i.e., what degree of artifact change is encompassed by the theory.
5) Testable propositions	Truth statements about the design theory.
6) Justificatory knowledge	The underlying knowledge or theory from the natural or social or design sciences that gives a basis and explanation for the design (kernel theories).
Additional components	
7) Principles of implementation (the causa efficiens)	A description of processes of implementing the theory (either product or method) in specific contexts.
8) Expository instantiation	A physical implementation of the artifact that can assist in representing the theory both as an expository device and for purposes of testing.

Table 9. Eight Components of an IS Design Theory

Source: Gregor and Jones (2007)

From a pragmatist perspective, the first component focuses on the purpose and scope of the developed artifacts related to the real-world problem that has to be solved. The constructs of interest are connected with the theoretical linkages that an artifact design should consider, such as the notion that designing a better user interface in terms of usability should result in a higher perceived ease of use. The third component should contain blueprints that explain how the artifact works—for example, how SPAT DPs work together. The fourth component suggests articulating the mutability of the developed artifact to make statements about how the artifacts may relate to existing theory. The fifth component is related to the outcomes of interest and proposes to derive testable propositions. This component focuses on constructs of interest and the effect of artifact design decisions that can be evaluated with an experiment—for example, the humanness of SPAT increases students’ learner engagement. The sixth component relates to kernel theories that serve as justificatory knowledge for the proposed effects when building and designing artifacts according to the theory. The last two components are additional and not necessary, but are often explicated in a design theory. On the

one hand this is about the principles of implementation that could be based on practical needs, the according implementation, and the derivation of more design guidelines and principles, and on the other hand it is about a physical implementation of the artifact (Gregor et al. 2020). DPs are used to specify design knowledge in an accessible and abstract form. Gregor et al. (2020) proposed a specific form to report DPs, which I used in my dissertation. The formulation of a DP includes a variety of components, as displayed in Table 10.

Title: DP Name	
Structure	Components
For <i>Implementer I</i> to achieve or allow <i>Aim A</i> for <i>User U</i>	Aim, Implementer, and User
In <i>Context C</i>	Context (Boundary conditions, implementation setting, further user characteristics)
Employ <i>Mechanisms M1, M2, M3 ...</i> Involving <i>Enactors E1, E2, E3, ...</i>	Mechanisms (acts, activities, processes, form/architecture, manipulation of other artifacts) Subsidiary components/artifacts that can have their own design principles
Because of <i>Rationale R</i>	Rationale Theoretical or empirical justification for the design principle

Table 10. Components of a DP

Source: Gregor et. al (2020)

As a demonstrative example, this study developed a DP called “conversational instruction” (DP2). This was as follows: For a SPAT Developer (Implementer I) to achieve or allow higher learning outcomes (Aim A) for students (User U) in online and blended learning environments, it is necessary to incorporate powerful machine learning frameworks (Mechanism M1) involving systematic text and speech recognition (Enactors E), because doing so allows SPAT to build a conversation with students similar to an educator–student interaction. The last part of a design theory involves strengthening the theory by an expository instantiation, particularly for testing the propositions derived. This also strengthens the design theory’s claim to solve actual problems and provide proof of the usefulness of the developed artifacts.

The knowledge contribution that is explicated in a design theory can be classified according to the framework developed by Gregor and Hevner (2013) considering

the dimensions of application domain maturity as well as solution maturity. The framework is presented in Figure 6.

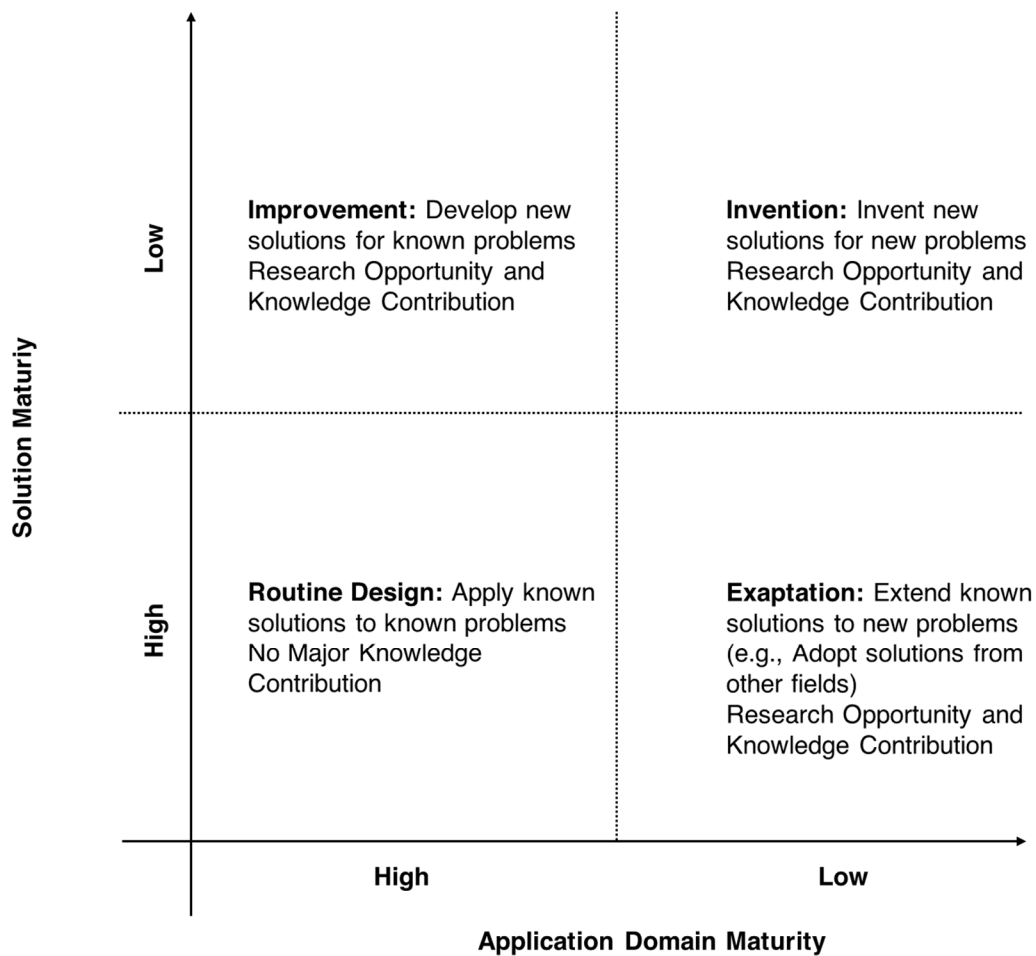


Figure 6. DSR Knowledge Contribution Framework

Source: Gregor and Hevner (2013)

A high majority of design projects fall into the category of high solution and application domain maturity. While these projects can be seen as routine designs without a significant design theoretical contribution, all other quadrants of the framework provide reasonable opportunity for design science researchers. This dissertation focuses on knowledge contributions in the area of improvement, meaning that it analyzes and develops new solutions (SPAT DPs and a learningbot-as-a-Service-Platform) for known problems (individual support in blended and online learning environments.)

Theories play a major role in DSR (Briggs 2006; Janson et al. 2017; Leimeister et al. 2005). Theories should inform design decisions that influence certain independent variables that are hypothesized to make a change in the DV under investigation. If these hypotheses are confirmed, (meta-)requirements of an artifact can be derived. For example, this dissertation hypothesizes that more interactive SPAT increases the DV “learning engagement” resulting in higher degrees of

“learning outcome”. The dissertation hypothesizes that the independent variable “scaffolding type” is a predictor of the outcome variable. I derived requirements from theory that relates to the outcome variable under investigation (see Section 4.2) and relevant DPs and features influencing the independent variable, as well as DV. Such an approach enables theoretically informed design decisions to be derived and helps researchers to better integrate the insights of the artifact evaluation into the greater body of knowledge (see section 7.1).

While this subsection has focused more on the role of theory as input and output, there are also process models that guide the design science research process helping to create design theories.

3.6.2 Conducting Design Science Research

Frameworks for conducting design science research help to guide the often fuzzy and iterative process. In IS research, two frameworks are preeminent. The first is the approach of Hevner and colleagues that was first formulated in Hevner et al. (2004) and complemented in Hevner (2007) through a cycle view of design science. This is depicted in Figure 7.

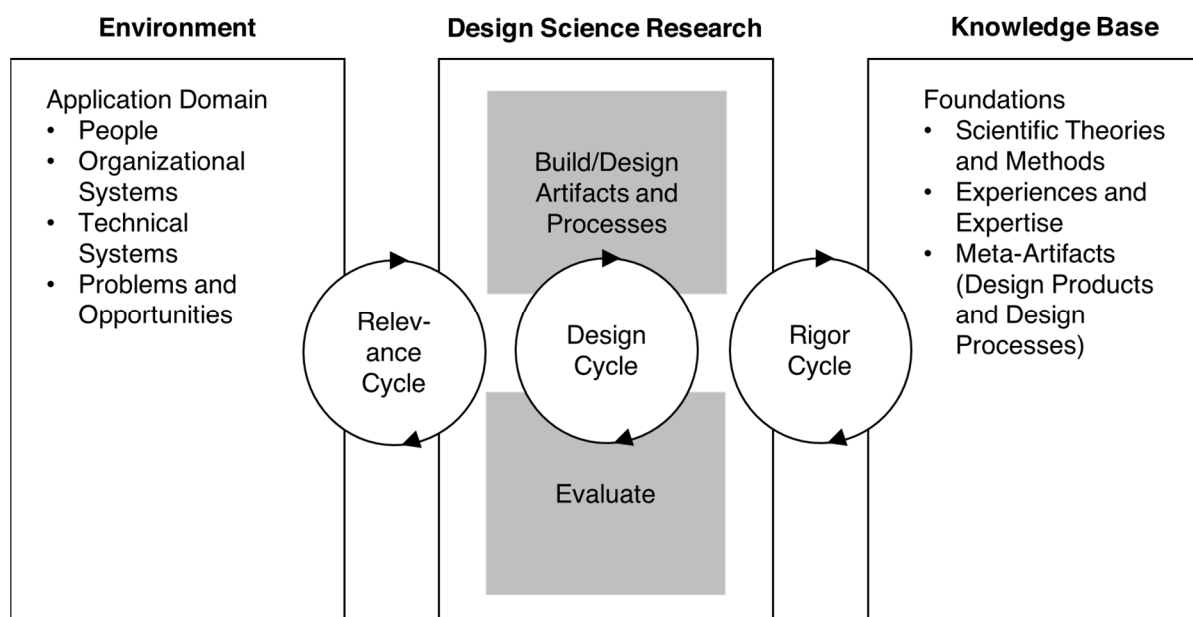


Figure 7. Design Science Research Cycles

Source: Hevner (2007)

The framework is divided into three cycles. The relevance cycle initiates a design-science project with a corresponding real problem from the environment, which is consistent with the pragmatic epistemological approach and links the environment to the design cycle (Hevner 2007). In the design cycle, the requirements of the stakeholders are taken into account, translated into a kind of design knowledge (e.g., DPs) and interactively tested with instantiated artifacts in real environments. Such testing in the field can lead to insights that inform additional

relevance cycles—for example, functional deficiencies or missing requirements that should be considered in the next design cycle. The rigor cycle refers to the role of theory and known methods during the design process. While recognized theories and methods ensure rigorous design through the whole process, additions to the knowledge base should connect to the experience already gained in design science research projects. Finally, the central cornerstone of a design science research project is the design cycle, which iterates between construction artifacts and evaluation, taking into account and, as discussed in Section 3.6.1, reinforcing the results of the relevance and rigor cycle to complete both cycles (Hevner 2007).

Another famous design science research approach focuses more on the process perspective and was developed by Peffers et al. (2007). The process is depicted in Figure 8.

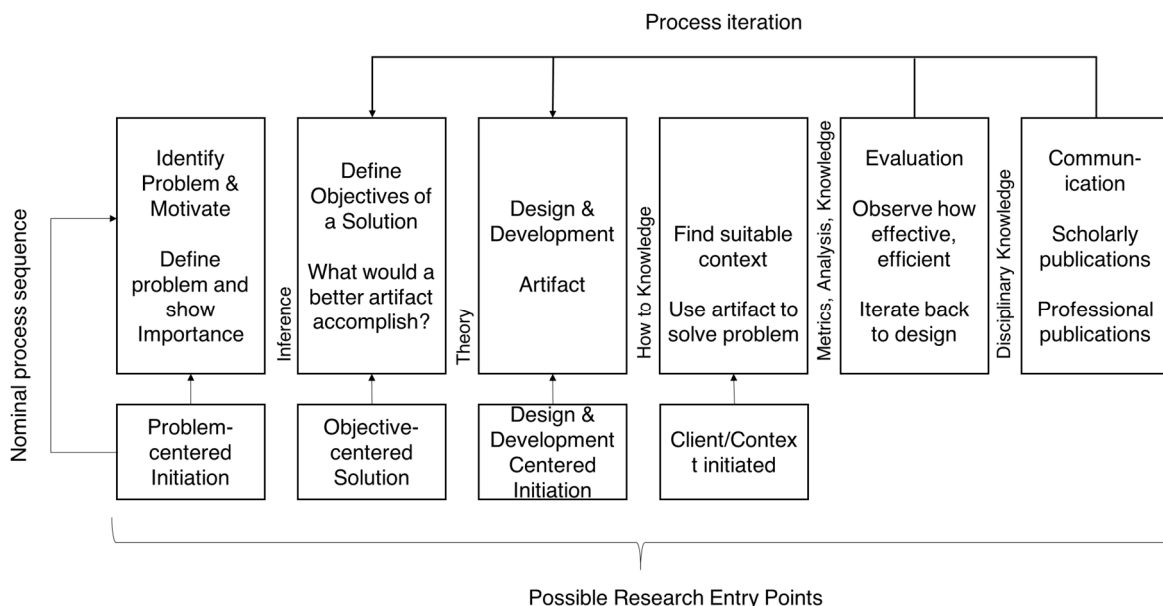


Figure 8. Design Science Research Process

Source: Peffers et al. (2007)

This design science research process consists of six major steps that need to be conducted. In the first step, and also in accordance with a pragmatic approach, the problem identification and motivation phase relates to the problem definition as well as the artifact's value justification. In the second step, the objectives for the solution are defined, representing the desired solution in quantitative or qualitative terms but also often as requirements that a solution should fulfill. The third step includes the design and development activities regarding the artifact, while the fourth step deals with demonstrating the usefulness of the solution instantiated in a particular context (experiments, for example). The fifth step describes the evaluation with, for example, pre-defined measures from the second step and could be embedded into a holistic evaluation framework (Venable et al. 2016). To

account for the iterative nature of a design process, researchers can iterate back to the second or third step to actively further develop the solution on the basis of the evaluation or proceed to the sixth step for communicating the research results.

This dissertation focuses mainly on the design science research approach of Hevner (2004), which emphasizes the interplay between relevance, design, and rigor cycle and focusses on the iterative nature of the development of the SPAT DPs and platform. This section concludes the philosophical and methodological background. In the further course of the dissertation, I answer the proposed research questions starting with a literature review to detect the current state-of-the-art of SPAT in education and a study to derive requirements and formulate DPs.

4 Developing a holistic Understanding of Smart Personal Assistant Technology in Learning Environments¹

In this section, I address and answer the first RQ of my dissertation:

RQ1: What are the requirements and relevant design principles of Smart Personal Assistant Technology in online and blended learning environments?

Answering this RQ lays the foundation of my dissertation by providing an overview of SPAT research in learning environments and highlights important research gaps that should be addressed. The RQ1 sub-questions seek to fill some of these gaps. First, in answering RQ1a I aim to describe the state of the art regarding SPAT in learning environments, and in RQ1b I derive more precise requirements for the development of SPAT in learning environments to formulate the first set of DPs.

¹ The insights presented in this section are partly based on two papers (Winkler and Söllner; Winkler and Roos (2019)). I thank my collaborators, the reviewers and attendees of the AoM 2018 and ICIS 2019 for their valuable feedback on my work.

4.1 Literature Review on the Current State of Smart Personal Assistant Technology in Learning Environments

4.1.1 Introduction

In this section of my dissertation, I address the first sub-question (RQ1a) of the first overarching RQ:

RQ1a: What is the state-of-the-art regarding Smart Personal Assistant Technology in learning environments?

The Section 4.1 provides insight into the state-of-the-art of SPAT from an input, process, and output perspective aligned to the TML framework developed by Gupta and Bostrom (2009) (see Section 2.1). By this means, I am able to identify important research gaps and provide recommendations for the next steps in SPAT research.

In recent decades, the number of students per lecturer has constantly risen (Nicol and Macfarlane-Dick 2006). Large-scale lectures at universities with more than 100 students per lecturer and open online courses (MOOCs) for large numbers of students are increasingly becoming the default learning environment. Consequently, individualized support provided by educators is nearly impossible, and students are unable to engage in effective learning (Brinton et al. 2015). Several studies have revealed that this lack of individualized support leads to weak learning outcomes, high dropout rates, and dissatisfaction (Brinton et al. 2015). The best solution would be to have one educator per student, but this is not possible due to financial and organizational restrictions (Oeste et al. 2015). SPAT has the potential to solve this problem using the examples of other sectors. SPAT has a growing presence in modern society, becoming an integral part of everything from personal assistants on mobile devices to technical support help over telephone lines, and even used for health interventions (Serban et al. 2017). In 2015, the size of the SPAT market was USD 113 million and is projected to be USD 994.5 million in 2024 (McKinsey & Company 2016). The usage of messaging apps has also increased exponentially in recent years.

Until now, knowledge about the effectiveness of SPAT in learning environments has been scarce. How this new technology affects the learning process of students and how students adopt these systems remains largely unexplored. Instead, studies have often tried to implement learning agents and more traditional types of intelligent tutoring systems in learning environments (Baker 2016; Goos et al. 1998; Govindasamy 2014). In contrast to these learning technologies, SPAT interacts with students in a synchronous way making it possible to react to individual intents. This allows students to remain in the driving seat and actively control their learning process, which is the major success factor for effective learning according to the predominant constructivist learning theory (Glaserfeld 1987).

The goal of this systematic literature review is to set the ground for future research regarding SPAT in learning environments. I therefore structured and synthesized

past research through the TML framework proposed by Gupta and Bostrom consisting of an input, process, and output perspective. This framework should help future researchers to fill important research gaps in order to understand the full potential of SPAT in learning environments. This paper is structured as follows.

4.1.2 Research Methodology

I conducted a systematic literature review based on the approach of Webster and Watson (2002) and vom Brocke (2015), according to which a review should be concept-centric and follow basic principles, such as defining the search scope and justifying all decisions made during the search process. First, I defined the search scope by using Cooper's Taxonomy (1988), as detailed in Table 13.

Charac- teristics	Categories			
Focus	Research Outcome	Research Methods	Theories	Practices or Applications
Goal	Integration	Criticism	Identifica- tion of cen- tral issues	
Perspec- tive	Neutral Representation		Espousal of Position	
Coverage	Exhaustive	Exhaustive with Selec- tive Citation	Representa- tive	Central or Pivotal
Organiza- tion	Historical	Conceptual		Methodological
Audience	Specialized Scholars	General Scholars	Practitioners or Policy Makers	General Public

Table 11. Framework for Defining the Search Scope

Source: Cooper (1988)

I focused on outcomes of SPAT research studies and their underlying theories and practical applications, because the major goal of the review is to show the potential of SPAT for increasing learning outcomes. Furthermore, I integrated and synthesized past literature by bridging the gap between management, information systems, education and psychology research with a common framework.

I used these three perspectives, because learning with a tutor is a very complex phenomenon investigated through different lenses by psychologists, educationists, computer scientists, and various others (Lajoie and Azevedo 2006). By this means, I was able to identify issues central to guiding future research directions. I reviewed SPAT literature from a neutral perspective, since the focus is to synthesize literature from different research fields and make an exhaustive literature review with selective citation. I tried to cover nearly the entire literature on SPAT in learning environments across different research areas such as education, psychology, and information systems and included a selected, representative sample of works in my findings. I tried to justify all my decisions and made the process transparent by following the five-stage framework suggested by Arksey and O'Malley (2005). The five stages of the SPAT literature review consisted of *a) identifying initial research questions; b) identifying relevant studies; c) making the study selection; d) charting the data; and e) collating, summarizing, and reporting the results.*

4.1.2.1 Step 1: Identifying the initial research questions

I analyzed and structured past literature from three perspectives aligned to the technology-mediated learning framework developed by Gupta and Bostrom (2009): input, process, and output. Based on that multi-perspective view, I derived the initial research questions (see Figure 9).

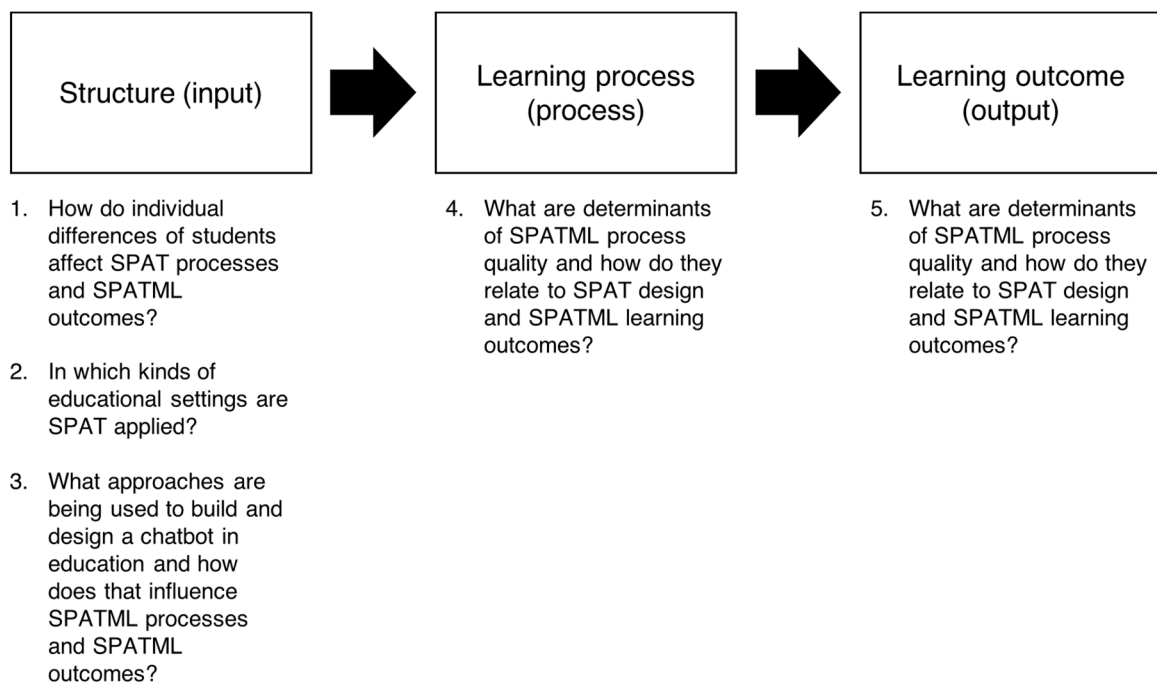


Figure 9. Initial Search Questions derived from our Framework

Source: Own illustration

First, I assumed an input perspective on SPAT in education consisting of individual student differences and different types of SPAT. I thus derived the following questions starting with the input perspective:

1. How do individual differences of students affect SPAT-mediated learning (SPATML) processes and SPATML outcomes?
2. In which learning environments is SPAT applied?
3. What approaches are being used to build and design SPAT in learning environments and how does that influence CML processes and CML outcomes?

Question one refers to student characteristics, which can have a significant impact on technology-mediated learning processes and outcomes (Gupta and Bostrom 2009). Question two tries to capture all fields of learning environments where SPAT is already successfully implemented, and question three highlights the importance of building and designing SPAT in order to have an impact on learning.

Second, I derived a question from the process perspective:

4. What are the determinants of SPATML process quality and how do they relate to SPAT design and SPATML learning outcomes?

Question four is a core question of this review, since even a well-developed SPAT might be useless if it does not produce additional value to students during the learning process.

Third, I derived an initial research question from the output perspective:

5. How can different kinds of learning outcomes be influenced by SPAT design, individual differences of students, and SPATML process quality?

To provide the basis for future research and practical implementation of SPAT, it is necessary to provide an overview of learning outcomes that can be achieved.

4.1.2.2 Step 2: Identifying relevant studies

I applied a wide definition of SPAT in order to capture available literature from different research fields such as education, psychology, and IS. First of all, I made an informal literature search in order to identify keywords for the actual open database search resulting in the search terms listed below:

("conversational agent" OR "chat bot" OR "Smart Personal Assistant Technology" OR "learning agent" OR "intelligent tutor system" OR "dialogue system" OR "smart personal assistant" OR "smart assistant")

After reviewing literature on the database *Science Direct*, I added "dialogue system" to the search string, because I considered this key word to be a central concept and judged that it would capture most of the SPAT literature. Before reviewing the literature, I developed inclusion and exclusion criteria. The last 10 to 15 years have seen a rapid improvement and spreading of SPAT due to new

technologies such as NLP, neural networks and machine learning (Pelk 2016). I therefore considered the time period 2005 to 2018 to be appropriate. I further included articles related to the fields of education, psychology, and IS. Moreover, I included studies in which SPAT was used to mediate any kind of learning process (i.e., where it helped to create learning outcomes of any kind.) I therefore excluded studies in which SPAT did not involve an intended learning outcome (e.g., customer support or just-for-fun SPAT), studies featuring bots with no chat function, and studies of recommendation and feedback systems with one-way communication (no multiturn dialog). Databases used for investigation were *Science Direct*, *EBSCO Business Source Premium*, *IEEE Xplore*, *Pro Quest*, *Aisle*, and *ACM DL*. In addition, I undertook a forward and backward search of central articles and used *Google Scholar* to identify other primary sources within grey literature.

4.1.2.3 Step 3: Study selection

Overall, I identified 1405 articles. A review of the abstracts revealed large numbers of articles that were irrelevant—for example, educational technologies that were not intended to communicate with students and only provided one-way recommendations and adaptations (Im and Hars 2007; Sonwalkar 2008). Furthermore, my literature storing system removed some articles because they were duplicates of articles on other databases. Guided by the defined inclusion and exclusion criteria, I identified 236 relevant articles. I thoroughly read the full texts for relevance, and this filtering process resulted in a representative and manageable total of 80 articles.

4.1.2.4 Step 4: Data charting and collation

In the next step, I mapped the concepts identified in the selected articles to the input-process-output framework and the initial research questions. For example, the article “Automated interventions for multiple health behaviors using conversational agents” by Bickmore et al. (2010) discusses the implementation and effect of SPAT in medicine, which addresses question two of our review. The next section discusses the findings of the review.

4.1.3 Results

The systematic review included 80 articles. Of these, 25 were published in 2018, 15 were published in 2017, and 12 were published in 2016. This reflects the trend of SPAT already described in Chapter 1. In the following subsections I structure and discuss my findings based on the initially developed research questions.

4.1.3.1 Question 1: How do students’ individual differences affect SPAT-mediated learning (SPATML) processes and SPATML outcomes?

Several individual differences in students influence SPATML process quality and the related learning outcomes through a complex interaction of students, learning methods, and structures (Gupta et al. 2010). First, the attitude of students toward

technology, and their trust in it, plays a major role. Students with a positive attitude to the value of SPAT feel more confident in their SPATML processes (Söllner et al. 2017). Second, learning characteristics such as trait emotions and personality traits have a significant effect on agent-directed emotions during the learning process. Students' trait emotions during learning—for example, anger, anxiety, or joy—and stable personality traits such as agreeableness, conscientiousness, and neuroticism, have a significant influence on SPAT-directed emotions during SPATML processes (Harley et al. 2012). Third, educational background and social and technological skills influence the way students interact with SPAT. For example, students with a background in humanities tend to have a more natural dialogue with human-like sentences during SPATML processes, whereas computer science students tend to be more formal and try to challenge SPAT to the limits of its “intelligence” (Novielli et al. 2010). Furthermore, students with better technological skills and a stronger need for interaction benefit more from SPAT (Ben Mimoun et al. 2015). Fourth, the self-efficacy and self-regulated skills of students influence SPATML processes and the related learning outcomes. High self-efficacy and well developed self-regulated skills in students have a positive effect on SPATML process quality and SPATML outcomes (Söllner et al. 2017). To summarize, the review shows that SPATML process quality and related learning outcomes depend significantly on individual student characteristics.

4.1.3.2 Question 2: In which learning environments is SPAT applied?

Health and well-being interventions

Several of the studies included in the review investigated SPAT used as a medical agent supporting students in their medical education or assisting patients with their therapy. In medical education, SPAT is often used on mobile devices in order to make learning independent of place and time. These examples of SPAT are mostly integrated into a learning platform, thus helping students through tests and instructional material (Alepis and Virvou 2011). In patient therapy, SPAT provides information and counseling to hospital patients at the time of hospital discharge and reacts to patients' questions. The aim of these interventions is to provide individual support to patients helping them to follow their therapy. For example, one study showed that patients with depressive symptoms prefer to receive support from SPAT rather than from doctors or nurses and showed that they have a significantly higher change to become healthy again compared to patients without access to SPAT. With the help of SPAT, patients learn how to help themselves with a degree of independence from human assistance (Bickmore et al. 2010). In the well-being sector, SPAT helps students to change their unhealthy behavior by triggering conversations with them. For example, daily conversations with SPAT can lead to greater physical activity and promote fruit and vegetable consumption (Bickmore et al. 2013).

Language learning

SPAT enables students to receive individual support in a mistake-friendly environment by giving them the chance to apply their language skills at any time. When provided with daily conversation contexts, students get the chance to use their knowledge in a simulated real-life situation. This reduces anxiety and increases the willingness to communicate in a foreign language (Ayedoun et al. 2015). Nevertheless, this effect seems to decline rapidly over time as students lose interest in SPAT as a language partner, more so than when students have human learning partners. This could be due to a simple novelty effect or may simply reflect the smaller uptake of SPAT compared to human assistants (Fryer et al. 2017).

Feedback and metacognitive thinking

SPAT in education can also be used to provide feedback to students about their performance and can trigger metacognitive thinking processes by discussing student learning behaviors in a synchronous way. One grade at the end of semester is often the only feedback students get, since lecturers have little time for individual formative feedback during the course. However, formative feedback during the learning processes is, according to Hattie (2012), one of the most important factors in increasing students' performance and motivation. SPAT used as a formative feedback tool has an additional positive effect on engagement indicators and task completion, since students can follow-up with questions on their assessments (Lundqvist et al. 2013). In addition, SPAT can also be used to help students improve their metacognitive thinking, which is considered a key competency in a rapidly changing world where lifelong learning becomes increasingly important. For example, in one study the use of SPAT during a course caused the discrepancy between computer assessment and student assessment to sink significantly (Kerly et al. 2008). The major advantage of SPAT in comparison to asynchronous ways of communication is that students receive real-time feedback to embed knowledge and assist their learning process during discussions.

Motivation and self-efficacy

SPAT can be used to increase student motivation in learning and strengthen individuals' self-perception. Motivation and self-efficacy are two of the most important factors of learning success (Schunk 1991). For example, SPAT can be used to create curiosity by asking students challenging questions, this being a main driver of intrinsic motivation (Oudeyer et al. 2016). Furthermore, SPAT can strengthen students' self-belief by giving them more control of their learning process (van der Meij et al. 2015).

Overall, the review reveals that SPAT is implemented for a variety of purposes. To summarize, SPAT tries to offer individual support in order to give people more control over their learning process. This leads to a higher CML process quality and higher learning outcomes.

4.1.3.3 Question 3: What approaches are being used to build and design a SPAT in learning settings and how does that influence SPATML processes and SPATML outcomes?

Flow-based SPAT

Flow SPAT is tree-based and uses a path previously defined by the SPAT's developer. A large number of articles in our review use this kind of SPAT. When a user input is entered, the SPAT tries to match it with the existing database entry. If a match is found, the answer is returned, if not, the SPAT tries to gain additional information or provides the user with buttons to choose from. This kind of SPAT requires a large dataset in order to minimize the frustration caused by wrong responses (Allison 2012).

Artificially Intelligent SPAT

Artificially intelligent SPAT allows the user to engage in a much freer way and is therefore more similar to a human-to-human conversation. In general, the literature differentiates between one-way and two-way artificially intelligent SPAT. In one-way artificially intelligent SPAT, the SPAT tries to understand what the user is saying by using machine learning techniques (Dutta 2017). In addition to this, two-way artificially intelligent SPAT uses artificial intelligence to feed the information back to the user. Instead of picking a response from a pool of pre-made answers, two-way artificially intelligent SPAT constructs the most accurate answer word-by-word before delivering it to the user. This type of SPAT is able to learn over time, based on real inputs from the users in terms of right and wrong answers given by SPAT. For example, the SPAT jabberwacky makes conversation by reflecting user input learned in similar contexts during previous conversations (Angeli and Brahnam 2008). Most of the recent studies in my review used one-way artificially intelligent SPAT to better understand student intentions and therefore increase SPATML process quality and learning outcome. It is worth noting that I was unable to find evidence of any two-way artificially intelligent SPAT in learning environments in studies performed prior to the end of 2018.

SPAT with integrated speech recognition

Some of the recent articles in the review investigate the benefits of speech input during learning sessions (Abdul-Kader et al. 2015; Griol et al. 2017; Saadatzi et al. 2017). This trend reflects the increasing interest in commercially available assistants such as Amazon's Alexa or Google's Home. Speech recognition within a SPAT converts user inputs by recognizing the vibrations in the air while the user is speaking. The SPAT then translates this analog data into digital data that is understandable to a computer (Graves and Jailty 2014). In addition to recognizing the content of speech messages, SPAT can also assess the emotional state of users and this can serve as further input for responses (Alepis and Virvou 2011). There is an opportunity here to identify affective behavior in students such as boredom, anxiety, excitement in order to create responses (Griol et al. 2017).

SPAT with integrated context-data

An increasing number of SPAT applications try to include contextual information to increase productivity and develop personality in the SPAT. Broadly speaking, the articles under review revealed two different ways of including context information in SPAT conversations. One method is to capture physical context data from students by analyzing their behavior on learning platforms. This enables SPAT to anticipate students' questions or adapt responses to their learning style leading to a higher SPATML process quality (Crockett et al. 2017; Latham et al. 2012; Latham et al. 2014). Another method of including context data in SPAT responses involves capturing the emotional status of students. Emotions such as anger, frustration, fear, and joy can be identified by speech-based emotion recognition as well as sentiment analysis of linguistic transcripts leading to more individualized learning support (Griol et al. 2017; Pérez et al. 2016).

My review has revealed that different types of SPAT have different influences on SPATML processes and the related learning outcomes. Identifying which kind of SPAT is most useful largely depends on the specific use case of SPAT in the learning process.

4.1.3.4 Question 4: What are determinants of SPATML process quality and how do they relate to SPAT design and SPATML learning outcomes?

Comparison between human-to-SPAT and human-to-human communication

The manner of communication in human-to-SPAT and human-to-human encounters differs significantly. Several studies in my review reveal that students interacting with SPAT during their learning process are less agreeable, less open, less extroverted, communicate for longer durations with shorter messages, lack much of the richness of vocabulary, and exhibit greater profanity (Corti and Gillespie 2016; Hill et al. 2015; Mou and Xu 2017). On the one hand, shorter messages and the lack of richness of vocabulary make it harder for SPAT to understand student' intents. On the other hand, people tend to adapt their language to the language of SPAT, in a similar manner to when they are speaking with children or using search engines such as Google (Hill et al. 2015). This can increase the effectiveness of SPAT. Moreover, the greater use of profanity by people when interacting with self-learning SPAT can lead to undesirable answers. An example of this is Microsoft's Tay SPAT on Twitter, which was manipulated to provide racist and profane responses over a 24-hour period (Vincent and James 2016). It is necessary for SPAT developers to understand this different mode of communication if they are to improve SPATML processes and learning outcomes.

Comparison between spoken and text input

One influencing factor of SPATML process quality is the input mode of SPAT. Some articles in my review investigated the comparison between text and spoken inputs, revealing that spoken input produces a richer language, a friendlier con-

versation, and longer dialogues (Cremonesi et al. 2017; Novielli et al. 2012). Furthermore, speech can be considered a more natural and convenient means of communication leading to increased user satisfaction. In addition, research in education shows that thinking aloud clearly improves a person's own thinking processes and recall capabilities (Bannert et al. 2017). Nevertheless, speech input is not superior to other input modes per se. There are learning environments such as lectures or bus journeys where speech input is not possible (Abdul-Kader et al. 2015). Hence, the combination of speech and input modes might lead to the best SPATML process quality.

Structural embeddedness in learning environment

Recent research has determined that the structural embeddedness of technology during the learning process is a key success factor for learning process quality (Bitzer and Janson; Gupta and Bostrom 2013). When students do not recognize the higher value of SPAT during learning immediately, and the instructions given on SPAT are not clear, the SPATML process quality is low. For example, if SPAT is not directly accessible where student questions arise, or if lecturers do not explain how to best use it, SPAT might remain unsuccessful (Söllner et al. 2017).

Despite different student characteristics and SPAT types classified as input factors, the manner in which students implement SPAT during their learning process is crucial to understanding its impact on learning outcomes.

4.1.3.5 Question 5: How can different kinds of learning outcomes be influenced by SPAT design, individual differences of students, and SPATML process quality?

Visualization of SPAT

The optic of SPAT is a key factor for achieving learning success in students. The articles under review drew contrasting conclusions. One article revealed that SPAT should be embodied, show emotions, and have a personality (Berry et al. 2005), while another was able to show via eye-tracking that an embodiment of SPAT is not needed, because students do not pay attention to it (Ben Mimoun et al. 2015). Moreover, the gender of SPAT does not play a role in measures such as SPAT usage and spoken topics (Brahnam and Angeli 2012).

Context-awareness

SPAT should include context information in order to increase cognitive and affective learning outcomes. This means that a SPAT should consider the cognitive as well as the emotional status of students in order to help them reach their learning goals. In particular, detecting knowledge gaps in students is essential for increasing their cognitive learning outcomes (Pérez et al. 2016). Moreover, detecting students' emotional status and including relational behavior such as caring and empathy helps students to improve their affective learning outcomes (Foster and

Oberlander 2010). This is also the reason why SPAT should include small talk (Bickmore et al. 2013; Kerly et al. 2008).

Proactive, guided by students

SPAT can help students to actively control and regulate learning processes in order to improve their own metacognitive learning outcomes. The predominant constructivist learning theory states that this is the only way that learning occurs (Glaserfeld, 1987). SPAT is therefore more successful when it empowers students to self-regulate their own learning process, leading to better metacognitive learning outcomes (Duffy and Azevedo 2015).

Integration in existing learning and instant messaging tools

SPAT can only be successful if it is easily accessible and responds fast. It should therefore be implemented in existing learning and messaging systems where students directly perform their tasks or communicate with each other. Some of the studies in my review showed that two of the main success factors of a SPAT are that students can receive the required information quickly, and they do not have to use too many applications during the learning process. The SPAT should therefore be implemented directly in the setting where students conduct tasks, such as learning management systems, or where they communicate outside of the learning platform, such as Facebook and Whatsapp (Pereira 2016).

The review showed that SPAT is able to improve affective, cognitive, and metacognitive learning outcomes. This makes it essential that SPAT is easily accessible, responds fast, empowers students to control their learning process, and considers contextual information such as the students' emotional status.

4.1.4 Discussion

In this section, I quickly summarize the findings, highlight and comment on important issues, and discuss the implications arising from my research questions.

4.1.4.1 Question 1: How do students' individual differences affect the SPAT-mediated learning processes and SPATML outcomes?

Several articles in my review showed that a number of individual differences in students, such as their personality traits, educational background, and self-regulated skills, have an impact on SPATML process quality. It therefore makes sense to tailor these systems to the individual differences of students. However, these findings should be interpreted with caution, since no study in the review empirically investigated whether personalized SPAT has a significant impact on learning outcomes. Studies that take these variables into account will need to be undertaken in the future. In addition to this, there are still many unanswered questions about how SPAT can be designed to adapt to individual differences such as personality traits and self-regulation skills.

4.1.4.2 Question 2: In which learning settings is SPAT applied?

Prior studies have noted the importance of SPAT for a variety of purposes including medical education and therapy, language learning, receiving feedback, and strengthening motivation and self-efficacy. Across all fields, it has been shown that students appreciate using SPAT due to a learning environment that is more mistake-friendly than one involving only human-to-human interaction. In a study by Bickmore et al. (2010), patients showed a preference for talking to SPAT rather than doctors and nurses. However, Fryer et al. (2017) stated that most of these effects occur because of a novelty effect and decline over time. This finding is consistent with technological implementations in other fields such as information systems (Wells et al. 2010). Hence, further work is required to establish the long-term effects of SPAT on learning processes and outcomes.

Very little was found in the literature on the question of systematically evaluating SPAT. This may be explained by the fact that learning processes and outcomes are complex and influenced by many factors. It is therefore difficult to preclude all factors other than SPAT. However, in the articles under review there are a variety of methods for evaluating the usefulness of SPAT. In therapy, studies used self-report ratings containing measures such as satisfaction and usability. In language learning, studies used metrics including the number of conversations (Ayedoun et al. 2015). In studies about receiving feedback and strengthening motivation and self-efficacy, researchers used dialogue time as an important metric (van der Meij et al. 2015). To develop a full picture of the potential of SPAT in education, additional studies regarding a systematic method for evaluating SPAT in education are necessary.

4.1.4.3 Question 3: What approaches are being used to build and design a SPAT in learning settings and how does that influence SPATML processes and SPATML outcomes?

My review showed that SPAT building approaches and SPAT design can be defined in three main dimensions: (1) whether they are retrieval based or generative; (2) whether they use speech or/and text as the input mode; and (3) whether they integrate context-data. Most of the studies in the review that were written before 2016 showed very basic, rule-based SPAT. With technological improvements regarding NLP and machine learning techniques, an increasing amount of generative, artificially intelligent SPAT arises. The most interesting finding was that of Dutta et al. (2017) who developed an intelligent SPAT tool to assist high school students in learning general knowledge subjects using NLP and machine learning algorithms for recognizing the correct intent. My review also showed that for closed domains with a fixed set of possible questions, rule-based SPAT might be more than sufficient. In all other cases, SPAT with artificial intelligence techniques leads to better SPATML process quality and SPATML outcomes. Hence, future research should concentrate on the development and design of artificially intelligent SPAT.

4.1.4.4 Question 4: How do students interact with SPAT during their learning process and how is that related to learning outcomes?

My review revealed that the way in which students interact with SPAT is fundamentally different to their interactions with human teaching staff. Studies show that students apply simpler sentences, lack rich vocabulary, and get distracted more easily. These findings suggest that for SPAT it is even more difficult to understand students' intents than for humans. The challenge here is to provide follow-up questions to gather the right intentions and to guide students back when the SPAT does not understand the intent or whenever students get distracted. This is an important issue for future research. Moreover, my review indicates that studies are not about replacing educators with SPAT. It is more about creating a co-dependent and intelligent relationship between educator and SPAT where the educator increasingly assumes the role of facilitator to deliver the best student experience (Hughes 2017). An important research question for the future is thus why and where human or SPAT assistance is preferable.

Furthermore, some of the recent studies under review recognized the success of voice assistants such as Amazon's Alexa and Google's Home and therefore included speech input in their educational SPAT. One interesting finding was made by Griol et al. (2017) who combined speech-based input and linguistic classification in order to successfully detect emotions. These findings support the work of other studies in this area that have attempted to include emotions in SPAT systems. Despite these promising results, questions remain as to how speech input changes SPATML learning processes and the resulting learning outcomes.

4.1.4.5 Question 5: How can different kinds of learning outcomes be influenced by SPAT design, individual differences of students, and SPATML process quality?

My findings reveal that SPAT can be used to improve different kinds of learning outcomes such as affective, cognitive, and metacognitive learning outcomes. However, I found contradictory results on how SPAT can achieve these learning outcomes. For example, the studies showed controversial insights regarding the visualization of SPAT. Some of the studies preferred to use embodiment and to build a personality for the SPAT (Berry et al. 2005), while others came to the conclusion that this lessens the SPAT's effectiveness (Ben Mimoun et al. 2015). Moreover, my findings suggest that SPAT should be used to empower students to actively control and regulate the learning process. Further studies should therefore analyze how SPAT can help students to actively learn new content and reflect on their learning process, and investigate which visualization method works best.

4.1.5 Implications for Research and Practice

In this section, I summarize and elaborate on unresolved issues with SPAT in education by using the input-process-output framework aligned to the TML frame-

work proposed by Gupta and Bostrom (2009), as shown in Table 12. Furthermore, I take a closer look at the potential of SPAT in education and outline practical implications.

	Structures (input)	Learning process (process)	Learning outcome (output)
Question 1	1. Adaptation of SPAT to individual student differences 2. Influence of personalized SPAT on learning process and outcome		
Question 2	3. Evaluation methods of SPAT in education		
Question 3	4. Design and development of artificially intelligent SPAT		
Question 4		5. Influence of conversational learning settings on constructing new knowledge 6. Integration of SPAT in different steps of the learning process with the help of learning theories 7. Human vs. SPAT assistance during the learning process	
Question 5			8. Empirical evidence on the influence of SPAT-mediated learning processes on learning outcomes 9. Influence of SPAT on self-efficacy and self-regulation skills 10. Influence of SPAT on long-term skill development

Table 12. Research Gaps structured by the Input-Process-Output Framework

Source: Own illustration

The literature review identified 10 important research gaps that need to be addressed in order to leverage the full potential of SPAT in education. The input perspective deals with different student characteristics and SPAT types influencing SPATML processes and SPATML outcomes. There are many unanswered

questions regarding how SPAT can successfully be tailored to students' individual differences and how personalized SPAT influences SPATML processes and SPATML outcomes. Moreover, the studies under review employed different ways of evaluating the usefulness of SPAT. Further investigation is needed to develop systematic evaluation methods for SPAT in education in order to distinguish good from bad SPAT. Furthermore, innovations such as Amazon's Alexa or Google's Home demonstrate how artificially intelligent SPAT can be successfully designed and developed for education. More research on this topic is required to provide greater insight into the design and development of educational SPAT.

From a process perspective, further research should be undertaken to investigate the additional value of conversations during the SPATML process in comparison to one-directional methods of learning such as learning materials or forums on a learning platform. This is an essential step for highlighting the superiority of SPAT over other learning technologies. Moreover, the inclusion of learning theories to identify parts of a learning process that can be supported by SPAT is crucial. Another identified research gap involves the important question of the point in the learning process at which human or SPAT assistance leads to better learning outcomes.

From an output perspective, more empirical evidence of SPATML processes on SPATML outcomes needs to be collected. Most studies have used a largely explorative research approach and have not utilized research designs that seek to measure significant increases in learning outcomes. Moreover, further work should be done to investigate how self-efficacy and self-regulation skills can be increased by SPAT. Furthermore, the existing literature does not address the effect of SPAT interactions on long-term skill development. It would be helpful to see how interactions with SPAT change students' skills in the long run.

This literature review shows practitioners the future potential and benefits of SPAT. Moreover, it helps to identify possible SPAT use cases for instructors that would offer additional, individual support to students during lectures.

4.1.6 Limitations

There were some limitations in my review method. I excluded promising fields SPAT application such as enterprise-customer interactions or just-for-fun SPAT, where learning is not a primary function. However, I believe that SPAT in education can learn from these fields. Moreover, my multi-perspective view of SPAT in education is only one way of collecting and synthesizing existing literature on SPAT. There are other technology-mediated learning frameworks that would probably have led to different results. Consequently, the suggested research gaps cannot be considered to be complete.

4.1.7 Conclusion and Next Steps for the Course of the Dissertation

The aim of this review was to synthesize, integrate, and structure existing SPAT literature in learning environments in order to set the basis for RQs 1b-3b. The

review has considered SPAT in learning environments from three different perspectives identifying crucial factors that influence the effectiveness of SPAT. The effectiveness of SPAT in education depends on individual student differences, the approaches to building SPAT, and the quality of the SPAT mediated learning process. Another aim of this review was to provide a framework for future research directions regarding SPAT in education. The review makes it evident that SPAT will play an increasingly important role in future learning environments. SPAT has the potential to create individual learning experiences for students and therefore increase learning outcomes and support lecturers and their teaching staff. This is especially true for large-scale lecturers and MOOCs with more than 100 students per lecture, where individual support is not possible due to financial and organizational restrictions.

In the next section of this dissertation, I use the insights about the state-of-the-art in SPAT and the formulated research gaps to derive more precise requirements regarding SPAT in learning environments. I complement these requirements suggested by theory with requirements arising from expert and student interviews, which together form the basis of the formulation of a first set of DPs for SPAT in learning environments.

4.2 Requirements of Theory and Practice to Design Smart Personal Assistant Technology

4.2.1 Introduction

In this subsection, I address and answer the second sub-question (RQ1b) of the first overarching RQ, which is related to the first challenge:

RQ1b: What are the relevant design principles of Smart Personal Assistant Technology?

Based on the findings of RQ1a, I rescanned the articles I had chosen and also conducted interviews to derive requirements and formulate a first set of DPs for SPAT in learning environments.

Currently, not much is known about how educators can design and integrate SPAT in their learning environments (Hobert and Wolff 2019). Past research in the field has mainly focused on specific implementations of SPAT and has failed to provide transferable insights and in-depth evaluations. Hence, the objective of this study is to formulate and validate a general set of DPs that researchers and practitioners can use to design effective SPAT as a means of individual learning support. I therefore follow the assertions of Hobert and Wolff (2019) and Kim and Baylor (2016) and state that there is a need for generalized design knowledge for SPAT in education.

To answer the stated research questions, I present two cycles of a DSR approach as described by Hevner (2007) (see Subsection 3.6.2.) In the first cycle, I derive requirements from the literature for SPAT in learning environments and conduct

qualitative interviews with educators and students. Based on this, I then formulate an initial set of DPs for SPAT. In the second cycle, I evaluate and refine the DPs.

4.2.2 Theoretical Background

To understand how SPAT can positively influence the learning process and outcome, I draw on the theoretical background of the ICAP framework, which is presented in Subsection 2.2.2. Each mode of the ICAP framework corresponds to different types of behaviors and knowledge-change processes predicting different learning outcomes. For example, when watching a video, students can view it without doing anything else (passive behavior). Students can also manipulate the recording by pausing, playing, fast-forwarding, and rewinding (active). A constructive behavior would involve self-explaining the concepts in the video (constructive). The most effective interactive student behavior would be if students discussed the content and its justifications with a peer or tutor (interactive). Properly designed SPAT has the capability to build dialog with students, bringing them from a passive to an interactive learning mode (Chi and Wylie 2014). To encourage interactive behavior in students, the interactions must be meaningful. Within meaningful interactions, both parties contribute a similar amount of input to the conversation and react to the answers of their discussion partner (Chi and Wylie 2014).

4.2.3 Research Methodology

I followed a DSR approach to answer the research questions. Specifically, I relied on Hevner's three cycle view to structure the research process (see Figure 10).

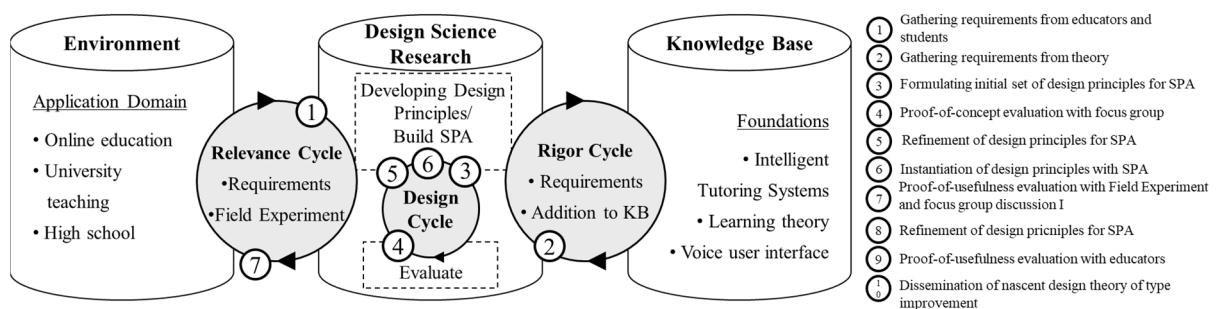


Figure 10. Design Science Research Process of the Paper

Source: based on Hevner (2007)

In step 1, I initiated the relevance cycle by gathering requirements from educator and student interviews. In step 2, I initiated the rigor cycle, gathering requirements from the literature review detailed in section 4.1 with three different theoretical perspectives: psychological, education, and IS. In step 3, I initiated the design cycle by formulating the initial set of DPs for SPAT in online and blended learning environments. In step 4, I conducted a first proof-of-concept evaluation with the help of a student focus group discussion. The objective of this evaluation was to make sure that I captured the most important requirements so that they could be logically translated into DPs (Sonnenberg and Vom Brocke 2012).

4.2.4 Designing Smart Personal Assistant Technology in Online and Blended Learning Environments

4.2.4.1 Step 1: Gathering Requirements from Students and Educators

My application domain involved students as well as educators from different levels and teaching formats (online education, university teaching, high school). As a first step, I wanted to establish a basic understanding of the current problems educators and students face in their everyday lives and to ascertain what they think of using SPAT to enrich existing online and blended learning environments. Hence, I conducted 20 interviews with students from high schools and universities and four interviews with educators. Table 1 shows the characteristics of the interview partners. All the interviews were semi-structured and lasted approximately 30 minutes each. I used a paper-based mock-up for showing interviewers our initial idea of SPAT and its possible use cases. The paper-based mock-up showed a possible interaction between a student and SPAT when working on a problem task. The problem task dealt with the issue of the rising incidence of traffic jams in a Swiss city. Table 13 shows the characteristics of the students and educators.

Inter-viewees	Characteristics
Students	Age: 21.35; level of education: 5 x high school, 10 x bachelor's, 5 x master's; nationality: 8 x Swiss, 11 x Germany, 1 x Italy; gender: 8 women, 12 men
Educators	Teaching experience: 5 to 25 years; type of school: 1 x high school teacher, 2 x bachelor's lecturer, 1 x master's lecturer; experience with different learning environments: 1 x online, 1 x mass lecture, 2 x small classes

Table 13. Sample Characteristics

Source: Own illustration

I created the interview questions with the research team in an internal workshop. In the first part of the student and educator interviews, I asked: What are the problems you encounter in a course? Can you explain these problems based on concrete examples? In the second part, I showed them a mock-up SPAT on a piece of paper demonstrating a student–SPAT learning dialogue. Based on this, I asked them about possible requirements for using this new technology. I transcribed the interviews and analyzed them using the method of user stories proposed by Cohn (2004). I identified USs in the interviews, and coded, clustered, and translated the USs into requirements for SPAT (see Table 14).

User Stories (Students)	Requirements from Practice (RP)
-------------------------	---------------------------------

USS1: As a <i>student</i> , I want the SPAT to give me immediate and detailed feedback on my responses.	RP1: SPAT should actively process and evaluate the students' responses.
USS2: As a <i>student</i> , I want the SPAT to reply as fast as a human in a conversation.	RP2: The SPAT platform should provide a fast hardware and software architecture.
USS3: As a <i>student</i> , I want to be able to use the SPAT at any time, in any place.	RP3: The SPAT should be available at all times.
USS4: As a <i>student</i> , I want to access the SPAT via voice as I would with a smartphone assistant (e.g., Google's Assistant or Apple's Siri).	RP4: The SPAT should include voice input and output.
User Stories (Educators)	Requirements from Practice (RP)
USE1: As an <i>educator</i> , I want to design SPAT by myself without needing a lot of technological knowledge.	RP5: The SPAT platform should allow educators to develop SPAT on their own.
USE2: As an <i>educator</i> , I want it to be possible for the SPAT to be used on different devices (e.g., smartphones, tablets, laptops).	RP6: The SPAT should be accessible via different devices.
USE3: As an <i>educator</i> , I want to use the SPAT for different learning goals and tasks.	RP7: The SPAT should be adaptable to different tasks in an efficient and uncomplicated way.
USE4: As an <i>educator</i> , I want to spend little or no money on developing SPAT for my courses.	RP8: The development costs of SPAT should be low.

Table 14. User-Stories and Requirements from Students and Educators

Source: Own illustration

4.2.4.2 Step 2: Gathering Requirements from Theory

I screened the articles I found in the literature review (see Section 4.1) and clustered similar requirements, which resulted in five requirement clusters. Table 15 depicts the perspectives, an exemplary paper, and the requirements from theory (RT).

Perspectives	Exemplary Paper	Requirements from Theory (RT)
--------------	-----------------	-------------------------------

Education	ICAP-Framework (Chi and Wylie 2014)	RT1: The SPAT should include a proactive and reactive interaction model, where the SPAT and the student both make a contribution.
	Scaffolding (Kim and Hannafin 2011)	RT2: The SPAT should provide different kinds of scaffolds during task completion.
Information Systems	Personification (Woolf et al. 2008)	RT3: The SPAT should include personification attributes.
	Error handling (Guo and Goh 2015)	RT4: The SPAT should resolve errors and dysfunctions so that the user can continue.
Psychology	Transparent Systems (Xu et al. 2014)	RT5: The SPAT should help the users to build an internal model of the functionality of the system.
	Engaging Response Structure (Pearl 2016)	RT6: The SPAT should use an engaging response structure different to text interfaces.

Table 15. Requirements from Theory

Source: Own illustration

4.2.4.3 Step 3: Formulating Design Principles Based on Requirements

Based on the identified requirements from theory and practice, I formulated an initial set of DPs aligned to the structure proposed by Gregor et al. (2020) (see Subsection 3.6.1). Figure 11 shows an overview of requirements, DP names, and design decisions. The rectangles in dark grey represent the requirements and DPs that I added in cycle 2 (after the first evaluation).

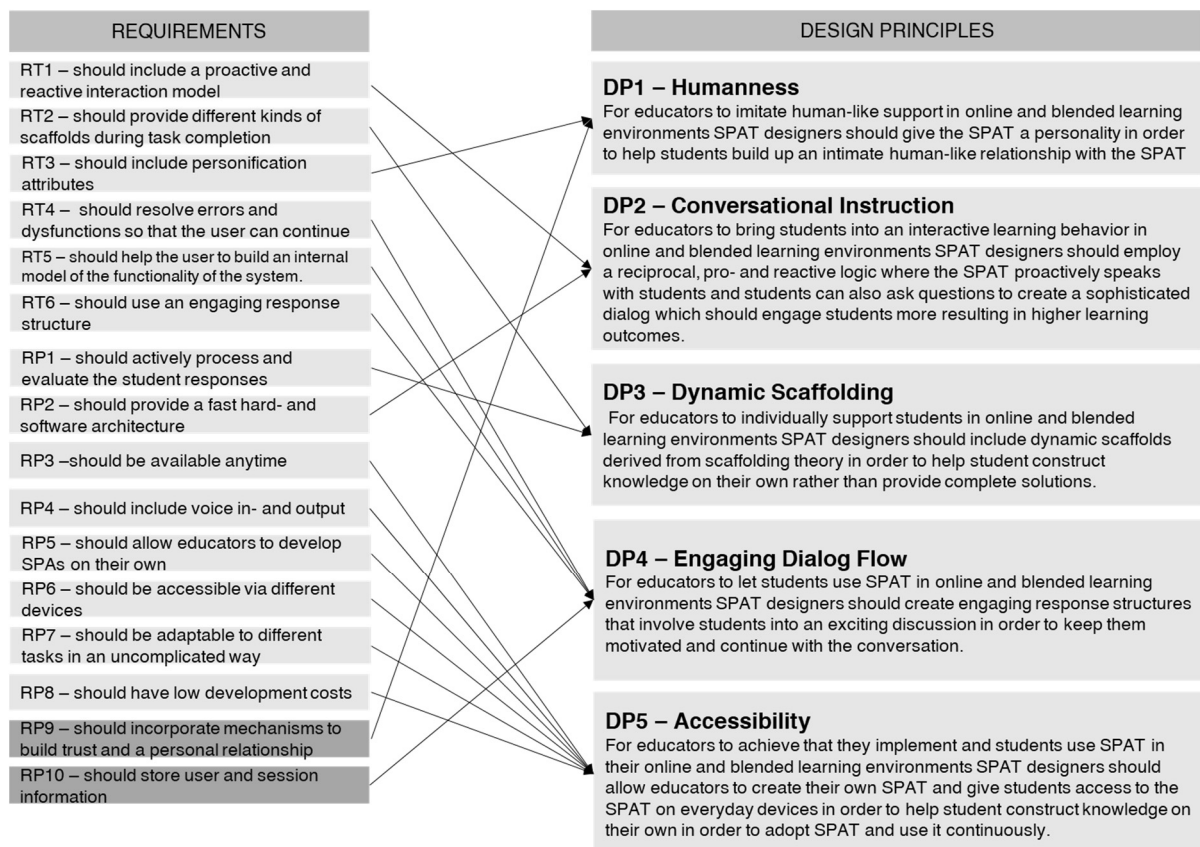


Figure 11. Requirements and Derivation of DPs

Source: Own illustration

Humanness (DP1) relates to how the SPAT is represented. For example, Woolf et al. (2008) revealed that personified computer artifacts show a higher user satisfaction and more intense usage. This DP is detailed in Table 16.

Design Principle Title	Humanness
Aim, Implementer, User	For educators to imitate human-like support
Context	In online and blended learning environments
Mechanism	SPAT designers should give the SPAT a personality
Rationale	To help students build an intimate human-like relationship with the SPAT

Table 16. DP 1 – Humanness

Source: Own illustration

Conversational Interactivity (DP2) is a fundamental principle of nearly every computer tutor. According to constructivist learning theories like the ICAP framework (Chi and Wylie 2014), interactivity is the main determinant for learning success. Hence, the SPAT tutor should try to trigger students' interactive learning behavior by using strategies to elicit constructivist responses from the student. DP2 is detailed in Table 17.

Design Principle Title	Conversational Instruction
Aim, Implementer, User	For educators to bring students into an interactive learning behavior
Context	In online and blended learning environments
Mechanism	SPAT designers should employ a reciprocal, proactive and reactive logic where the SPAT proactively speaks with students and students can also ask questions
Rationale	To create a sophisticated dialog which should engage students more resulting in higher learning outcomes

Table 17. DP 2 – Conversational Instruction

Source: Own illustration

Dynamic Scaffolding for complex tasks (DP3) is the second main principle coming from educational research. Scaffolding is an effective standard tutoring method. It nudges students toward acquiring skills on their own and is applicable in almost every learning environment (see Table 18).

Design Principle Title	Dynamic Scaffolding
Aim, Implementer, User	For educators to individually support students
Context	In online and blended learning environments
Mechanism	SPAT designers should include dynamic scaffolds derived from scaffolding theory
Rationale	To help students construct knowledge on their own rather than provide complete solutions

Table 18. DP 3 – Dynamic Scaffolding

Source: Own illustration

For educators to individually support students in online and blended learning environments, SPAT designers should include dynamic scaffolds derived from scaffolding theory in order to help students construct knowledge on their own. En-

agement (DP4) is the fourth DP. Similar to human-to-human conversations, dialogs must be engaging so that both partners are motivated to contribute to the discussion. Looking at voice interaction strategies, it becomes clear that a major goal and challenge is to keep users engaged during the conversation (Pearl 2016). For educators to let students use SPAT, SPAT should create engaging response structures that involve students in a stimulating discussion (see Table 19).

Design Principle Title	Engaging Dialog Flow
Aim, Implementer, User	For educators to let students use SPAT
Context	In online and blended learning environments
Mechanism	SPAT designers should create engaging response structures that involve students in a stimulating discussion
Rationale	To keep students motivated and continue the conversation

Table 19. DP 4 – Engaging Dialog Flow

Source: Own illustration

A key reason why ITS is not more widespread is the high cost and time effort involved in changing learning contexts. By enabling educators to create their own SPAT, the costs and effort can be significantly reduced while individualized support can be increased (see Table 20).

Design Principle Title	Accessibility
Aim, Implementer, User	For educators to achieve implementation of SPAT and for students to use it
Context	In their online and blended learning environments
Mechanism	SPAT designers should allow educators to create their own SPAT and give students access to the SPAT on everyday devices
Rationale	To help educators and students adopt SPAT and use it continuously

Table 20. DP 5 – Accessibility

Source: Own illustration

For educators to adopt SPAT and use it in their online and blended learning environments, SPAT designers should allow educators to create their own SPAT and give students access to the SPAT on the devices they use.

4.2.4.4 Steps 4 and 5: Proof-of-Concept Evaluation and Refinement of DPs

According to Sonnenberg and vom Brocke (2012), it is important to focus on two aspects during evaluations: (1) the constituents of the artifact and the design decisions taken; and (2) the evaluation of the usefulness of the artifact. The first evaluation concentrates on the constituents of the artifact. Focus group discussions are considered a suitable method for evaluating designs within DSR projects (Sonnenberg and Vom Brocke 2012; Venable and Baskerville 2012). In my study, the goal of the focus group discussion was to check the validity of requirements and translate them to the resulting DPs. Moreover, I wanted to gather further requirements and DPs for the next cycle. The focus group discussion lasted 60 minutes with one of the researchers as facilitator. The participants were six master's and four undergraduate students with an average age of 23.5 ($m=3$, $w=7$), and they were collected from a Swiss business university (with different students to those used in step 1). The focus group discussion was structured as follows. In the first part, one of the researchers introduced the goal of the group discussion. In the second part, I showed students the derived requirements and DPs and asked them to rate the requirements according to their importance and check for weaknesses regarding the formulation of DPs. In the third part, I used a small Wizard-of-Oz Simulation where one of the researchers (not the facilitator) was located in a different room using an Alexa Echo Dot and its intercom feature to talk to the students. I also placed an Alexa Echo Dot in the room the students were in so that the researcher was able to guide them through a 10-minute learning task. After the short simulation, I asked students to collect further requirements and DPs. I included the additional requirements and DPs in the overview given in Figure 2 (see dark grey elements).

4.2.5 Conclusion and Next Steps for the Course of the Dissertation

In this study, the goal was to derive requirements of SPAT in learning environments and formulate a first set of DPs. Based on the findings of RQ1, I scanned the articles again and searched for requirements in theory. I enriched this data with interviews with experts and students. As SPAT will increasingly enter students' everyday lives, it is important to understand how SPAT should be designed in online and blended learning environments. Regarding practical implications, I provide guidelines on how to build SPAT for students' online and blended learning environments. The formulated DPs are the starting point for the next overarching research question, where I test the most relevant DPs in a series of different laboratory and field experiments in online and blended learning environments. To summarize, the most relevant DPs I discovered were humanness, conversational instruction, dynamic scaffolding, engaging dialog flow, and accessibility.

5 Design of Smart Personal Assistant Technology in Online and Blended Learning Environments²

RQ1a helped to identify current research needs and to align the dissertation with the current state of research. In RQ1b, I went into more detail and looked at the normative requirements for SPAT in online and blended learning environments. These requirements resulted in five relevant DPs. The overarching RQ2 acknowledges these requirements and investigates how these DPs influence students' learning processes and corresponding learning outcomes at different levels of education, in different learning environments (online and blended learning), and for different kinds of outcomes. The overarching RQ is as follows:

RQ2: How Smart Personal Assistant Technology be designed to improve students' learning processes and learning outcomes in online and blended learning environments?

RQ2 is divided into two sub-questions. RQ2a investigates two SPAT DPs derived from RQ1 (humanness and conversational instruction) in blended learning environments, while RQ2b investigates two SPAT DPs (dynamic scaffolding and engaging dialog flow) in pure online learning environments. This approach helped me to evaluate the influence of the DPs in different settings and thereby strengthen their generalizability.

5.1 Experimental Evaluation of Smart Personal Assistant Technology to Improve Group Task Outcome

5.1.1 Introduction

In this subsection I address RQ2a, which deals with the influence of humanness and conversational instruction (DP1 and DP2) in SPAT on learning processes and outcomes within a collaborative blended learning environment, which is relevant to many current higher education courses. This paper forms one of two studies that help to answer RQ2a:

How can humanness and conversational instruction incorporated into Smart Personal Assistant Technology improve students' learning success in blended learning environments?

The starting point for the answer to this RQ is the formulated DPs and particularly DP 1, humanness. I translated this DP into DFs and integrated these into a SPAT.

Effective collaboration between group participants does not happen naturally (Briggs et al. 2003). It is a complicated mechanism that needs guidelines in order to be effective (Briggs et al. 2013). Past literature demonstrates that human facilitators are able to improve group performance (Anson et al. 1995; Oxley et al.

² The insights presented in this section are partly based on various papers (Winkler et al. (2019b); Winkler et al. (2019a); Winkler et al. (2019c); Winkler et al. (2020b)). I thank my collaborators, the anonymous reviewers and attendees of the WI 2019, AoM 2020, CHI2020, and the senior editor and anonymous reviewers of the Computers and Education Journal for their valuable feedback on my work.

1996; Witte 2007), and some studies indicate that new emerging SPAT might also be able to compete with human facilitators or even outperform them in some specific contexts (Seeber et al. 2018; Vanlehn 2011). For example, Lucas et al. (2014) found that group members are more willing to disclose when they interact with a SPAT than when they interact with a human. Interacting with a computer tutor lowers evaluation fears and impression management (Lucas et al. 2014). Moreover, interacting with SPAT might also lead to a higher degree of agency in the group. The degree of agency is a key factor in how people experience interactions with technology and humans (Limerick et al. 2014). Although the current state of technology is not yet able to keep up with expert human facilitators, recent technological developments in machine learning and natural language processing (NLP) suggest that SPAT will soon be able to imitate human facilitators (Seeber et al. 2018). A better understanding of how newly emerging SPAT can increase students' learning outcomes and replace costly human tutors in future use scenarios will help to inform future designs of SPAT for student group support.

To partially answer the research question as one of two papers, I based the work on research in the field of computer-supported cooperative work (CSCW) and collaborative problem-solving (CPS). Based on this, I conducted a laboratory experiment with state-of-the-art SPAT support. Using a mixed-method approach, I complemented the quantitative results (RQ1) with a qualitative CA approach (RQ2). Mixed-method research can help develop rich insights into various phenomena of interest that cannot be fully understood using only a quantitative or a qualitative method. Currently, there is a dearth of mixed method research in IS (Venkatesh et al. 2013). In particular, CA approaches have seldom been used in IS studies (Sarker et al. 2013).

In my experiment, groups of three participants had to conduct a 30-minute problem task, receiving help from either SPAT or a scripted human facilitator. The scripted human facilitator provided exactly the same scope of information and functions as those incorporated into the SPAT facilitator. I decided to limit the human facilitators' input to a script, because this study is one of the first to make direct comparisons between SPAT and human facilitators and is intended to serve as a basis for answering further questions on the extent to which SPAT tutors can be as effective as human facilitators. By scripting the human facilitator, I was able to control other potential factors (prior knowledge of the human facilitator, etc.). The results showed that SPAT facilitators outperform scripted human facilitators and produced three interaction patterns in which human–SPAT interaction differs from human–human interaction. The results suggest that SPAT might be a suitable replacement for human facilitators and might even show some additional benefits. The next subsection classifies the term SPAT, and describes computer-supported collaborative work and CPS. Section 5.3 describes the setup of the lab experiment, including sample description, task design, technological

development of the SPAT, and the measurement details. Finally, I report our results, discuss them, and close by mentioning limitations, offering a future outlook and a conclusion and suggesting directions for future research.

5.1.2 Theoretical Background and Hypotheses Development

In this section, I describe the CSCW and collaborative problem-solving that serves as a foundation for my hypotheses.

5.1.2.1 Computer-Supported Cooperative Work

Computer-supported cooperative work CSCW aims to improve the productivity and effectiveness of multiple individuals working together in a conscious way (Mills 2003). CSCW consists of software tools and technology that support a group of individuals working together and adopts either a technology-centric or work-centric viewpoint. A technology-centric viewpoint emphasizes designing computer technology to support groups working together. A work-centric viewpoint emphasizes designing computer systems to support group work (Grudin 1994). Among others, there are four important dimensions inherent in CSCW (Mills 2003). One important dimension of CSCW technology occurs on the time dimension: is there a requirement to support cooperative work that occurs simultaneously (synchronously) or separately (asynchronously) or both? Another decision relates to space: must the individual collaborators be physically located at the same site, such as a room or an auditorium? A third important dimension is group size: must the system support a small team or a whole department? A fourth dimension involves the interaction style: does the group require support for planned or spontaneous interactions or both? Our work concentrates on simultaneous cooperative work, where collaborators are physically located in the same room, because this kind of group work is one of the most common collaborative working environments in practice and allows interfering variables within a laboratory experiment to be controlled. Moreover, I concentrate on small groups where interactions are taking place spontaneously and are driven by the participants themselves.

5.1.2.2 Collaborative problem-solving

Collaboration is used to describe a variety of behaviors and generally occurs when more than one person works on a single task toward a common goal (Hesse et al. 2015), and group members “must make a conscious, continued effort to coordinate their language and activity with respect to shared knowledge” (Roschelle and Teasley 1995). CPS is a specific form of collaboration, namely the responsive approach of a team when jointly solving a problem by exchanging ideas (Hesse et al. 2015). Modern workplaces demand people who can solve problems together with others (Nelson 1999; PISA 2017). Effective CPS does not happen naturally. CPS is a complex process that can be defined as “a joint activity where two or more people work together to progress through a series of cognitive states to solve a problem”. CPS incorporates the components of cognition found in individual problem-solving (cognitive dimension) in addition to the components

of collaboration (social dimension) (Hesse et al. 2015, Dillenbourg 1999). The cognitive component consists of taking appropriate actions to solve the problem (i.e., it encompasses the team members' ability to define problems, an understanding of their prior knowledge, the identification of gaps in what is known and what is required, the integration of these in order to generate solutions, and the monitoring of the group process toward achieving the common goals (Graesser et al. 2018b)), and the social component consists of maintaining a shared understanding across the group members (i.e., the coordination of all participants' actions (Hesse et al. 2015; PISA 2017)). Shared knowledge building often resonates with common ground and mental models of a team (Roschelle 2013). In this context, the facilitator's role is to provide ongoing and individualized feedback to the group members to improve their cognitive and social actions (Merrill et al. 1996). Computers have often been employed in an effort to support groups in problem-solving activities (Jeong and Hmelo-Silver 2016). For example, Harsley et al. (2016) found that student groups were able to significantly increase their programming skills while receiving help from a SPAT facilitator within a CPS environment. Similarly, Mcalister (2004) demonstrated that group members interacting with SPAT-similar systems showed higher argumentation skills leading to better task outcomes. Tarouco et al. (2018) logged and analyzed student-SPAT dialogs within a CPS environment and concluded that students tend to perceive SPAT as their collaboration partner.

New emerging SPAT has the capacity to track the contributions of team members, provide feedback automatically, and offer recommendations to team members to guide them through a task (Graesser et al. 2017a). Some studies indicate that new emerging SPAT might also be able to compete with human facilitators to some extent or even outperform them in specific contexts (Seeber et al. 2018; Vanlehn 2011). For example, Lucas et al. (2014) found that group members are more willing to disclose when interacting with SPAT than they are when interacting with a human. Interacting with a computer lowers evaluation fears and impression management (Lucas et al. 2014). One reason for this might be that group members could easily feel watched or distracted when a facilitator, who will often be a stranger, tries to advise them (Xolocotzin Eligio et al. 2012). Furthermore, interacting with SPAT might lead to a higher degree of agency being perceived by the group members. The degree of agency is a key factor in how people experience interactions with technology and humans (Limerick et al. 2014). Additionally, recent studies in the area of trust indicate that people tend to trust automated machines more than humans in specific conditions (Yang et al. 2017). Furthermore, recent developments in voice recognition and machine learning have turned SPAT into a powerful interaction partner similar to human partners (Seeber et al. 2018). Nevertheless, the current state of technology is not yet able to keep up with expert human facilitators, but given that SPAT and human facilitators are limited to the same scope of information and functions, I assume that groups interacting with SPAT facilitators are able to achieve better task group outcomes. Hence, I propose:

H1. Groups interacting with human-like SPAT facilitators achieve better group task outcomes compared to groups interacting with scripted human facilitators within a problem-solving context.

Additionally, SPAT facilitators might also have different effects on the groups' collaboration quality itself. Past research indicates that SPAT and human facilitators have a different effect on participants' emotions (Litman et al. 2004). For example, Vail et al. (2016) found that human facilitators can have a significant impact on participants' affective response, such as joy or frustration. Moreover, Pickard et al. (2016) were able to show that people tend to be more open to SPAT, and that they feel more comfortable than without this and are more likely to make negative self-admissions when using the technology. Particularly in relationships of authority, where an often unknown facilitator provides advice, participants might feel controlled and are therefore easily inhibited in their collaborative behavior. Furthermore, Gratch et al. (2007) found that virtual humans can be more engaging than real ones, an outcome that was measured by the total number of words participants spoke during an interaction. Given that SPAT and human facilitators are limited to the same scope of information and functions, I assume that SPAT facilitators are able to increase the collaboration quality within the group. Hence, I propose:

H2. Groups interacting with human-like SPAT achieve higher levels of collaboration quality compared to groups interacting with scripted human facilitators within a problem-solving context.

5.1.3 Research Methodology

To test the hypotheses, I conducted a laboratory experiment in which groups of three solved a 30-minute problem task while receiving help from either SPAT or scripted human facilitators. The main goal was to investigate whether the human-ness DP incorporated in SPAT could keep up with human tutors. The treatment group (TG) was able to communicate with the SPAT facilitator via an Amazon Alexa Echo Dot device, while the control group was able to communicate with the scripted human facilitator located in the same room. I use a scripted human facilitator as a baseline for two reasons. First, the current state of technology is not yet able to keep up with expert human facilitators. Second, I wanted to ensure that the two groups were as comparable as possible in order to show differences between interactions with SPAT and human facilitators. The human facilitator followed a script with exactly the same scope of information and functions as those incorporated into the SPAT. The groups were not allowed to use the internet or any other aid.

5.1.3.1 Sample

The sample consisted of 127 participants. I had to exclude seven participants from the sample because they did not show up, resulting in subgroups of fewer than three. Altogether, I had 20 groups of three in the TG (SPAT) and 20 groups

of three in the control group (human). The participants were students from a Swiss business university lab pool with an average age of 21.59 (undergraduate = 82, graduate = 38). Most of the students were studying business-related subjects (e.g., economics, business studies, law). The sample consisted of 68 men and 52 women. All participants were randomly assigned to one of the subgroups. Table 21 shows the characteristics of the two groups.

	Treatment Group (SPAT)	Control Group (Human)
Gender	m = 35, w = 25	m = 33, w = 27
Average Age	21.29	21.88
Study background	18 x economics, 30 x business studies, 10 x law, 2 x others	15 x economics, 31 x business studies, 9 x law, 5 x others
Pre-experience with SPAT	5.21 (out of 7)	5.02 (out of 7)
Personal innovativeness	4.83 (out of 7)	4.79 (out of 7)

Table 21. Sample Characteristics

Source: Own illustration

I conducted ANOVA tests to ensure that the two groups were equal in terms of gender, age, study background, pre-experience with SPAT, and personal innovativeness. The randomization was successful, since all results are not significant ($p < 0.05$). Participants received a baseline monetary reward for participating with an optional additional reward depending on their group task outcome.

5.1.3.2 Task Design

The goal of the study was to identify whether human-like SPAT facilitators are better able to increase group task outcomes and the quality of collaboration compared to scripted human facilitators within a problem-solving context. Hence, it was important to choose a well-defined problem task. Problem-based learning theory describes a task as a problem task when it is ill-structured, open-ended, realistic, and resonates with the executors' experience (Jonassen 2011). A problem task definition should further promote conjecture and argumentation (Hmelo-Silver 2004). Based on the described criteria of problem task definitions, I decided to adapt a problem task from Gallagher et al. (1992). Table 2 shows how the task

addresses the criteria of a problem task as defined by Jonassen (2011) and Hmelo-Silver (2004).

Task: Increasing traffic flow has led to a significant increase in the number of traffic congestions, accidents, and road casualties in Zürich. The city of Zürich asks you to build a focus group and find the best solution for them. What should the city do about it?	
Requirements of a problem task definition	Description of how our task addresses the requirements
R1: Ill-structured	The task can be considered as ill-structured because not every piece of information is given. Moreover, it gives no advice on how to solve the task.
R2: Open-ended	The task allows different solutions to be developed. “The right solution” does not exist.
R3: Realistic	This is a problem that cities frequently have to deal with, since increasing traffic flow is a fact. It can therefore be considered as realistic.
R4: Resonate with the executors’ experience	Since every participant has been in a car before and most of them have a driver’s license, they can empathize with the problem task.
R5: Promote conjecture and argumentation	The task is missing information about the causes, the core of the problem, etc. It therefore opens room for conjecture and different kinds of argumentations.

Table 22. Group Task and Requirements

Source: Own illustration

5.1.3.3 Technological Development of the Smart Personal Assistant Technology

To answer the research questions and show the intended effects, it was necessary to develop a human-like SPAT that is able to help the group members in the best possible way. According to Doyle et al. (2019), humanness needs to be considered not as unidimensional, but as a multi-dimensional concept in speech interface design. I incorporated two major humanness design features (DFs). First, I addressed voice quality, which is a major DF of humanness. I used Amazon’s Alexa Voice Services (AVS) and Skill Development Kit 2.0 with nodeJS, since this programming framework offers the most human-like voice and is the current state-of-the-art in speech recognition and NLP. Alexa modulates her voice and

makes human-like pauses while she speaks. Second, I incorporated small-talk and some conversational storytelling similar to the expected behavior of a human tutor (see Figure 12). In human conversations, it is customary to begin with small talk before “getting down to business”. The purpose of this small talk is primarily to build rapport and trust among the conversation partners (Bickmore et al. 1999). Conversational storytelling should be an integral part of an agent’s small talk capability (Cheepen 1988).

I included a small talk passage at the beginning of the conversation, making Alexa ask the students how they were doing and using the small talk built-in function of Alexa Voice Services, which enables students to ask questions such as “Alexa, are you male or female?”. Moreover, we included storytelling by introducing Alexa as the students’ learning partner and by explaining the different problem-solving steps and how Alexa uses problem-solving steps derived from theory to guide students through the task. Figure 12 shows an excerpt of an exemplary group dialogue with the SPAT facilitator and the corresponding DFs (the communication between group members is not depicted).

Furthermore, I used the Alexa Skill Development Kit 2.0, which provides programmatic access to Alexa features along with several tutorials and blueprints, so I was able to use basic coding patterns to create rich, interactive experiences (Amazon 2019).

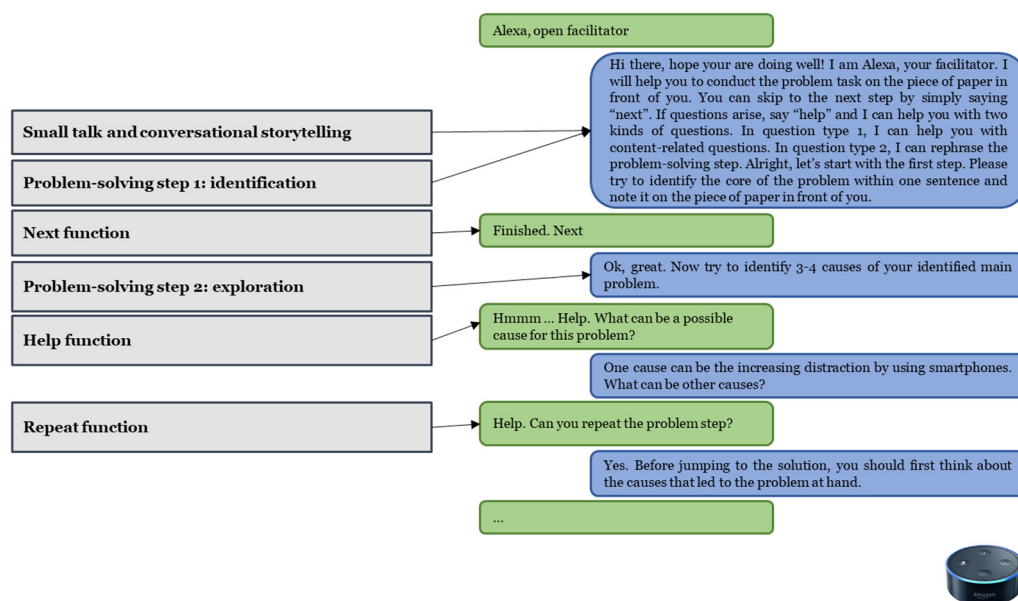


Figure 12. SPAT DFs and Excerpt of Example Dialog with SPAT

Source: Own illustration

During the conversation, the SPAT facilitator proactively guided students through the task by providing five different problem-solving steps adapted from Kim and Hannafin’s (2011) problem-solving phases: (a) *identification (define the problem)*;

(b) exploration (search information) (c) reconstruction (select and process information); (d) presentation and communication (information aggregation and communicate result); and (e) reflection and negotiation (looking back, revising ideas and theories). I thus ensured that the SPAT facilitator was able to help participants to solve the problem in a structured way. The SPAT facilitator provided its problem-solving steps at intervals of five minutes. The participants were also able to skip to the next step if they had finished the previous step before the end of the time provided. The SPAT was also able to provide hints during the steps. The participants could either ask for content-specific help or ask the facilitator to rephrase the current problem-solving step.

5.1.3.4 Experimental Procedure

Participants in the TG (SPAT facilitator group) received the problem task on a piece of paper and were able to access the facilitator with the help of a standalone SPAT device (Amazon's Alexa Echo Dot). Participants in the control group (scripted human facilitator) received the problem task on a piece of paper and the human facilitator was located in the same room as the participant group using a script with exactly the same scope of information and functions as those incorporated in Alexa. As already mentioned in subsection 5.1.1, I scripted the human facilitator because I wanted to control for other potential factors (such as prior knowledge of the human facilitator, etc.). See Figure 13 for an exemplary set up involving both TGs.

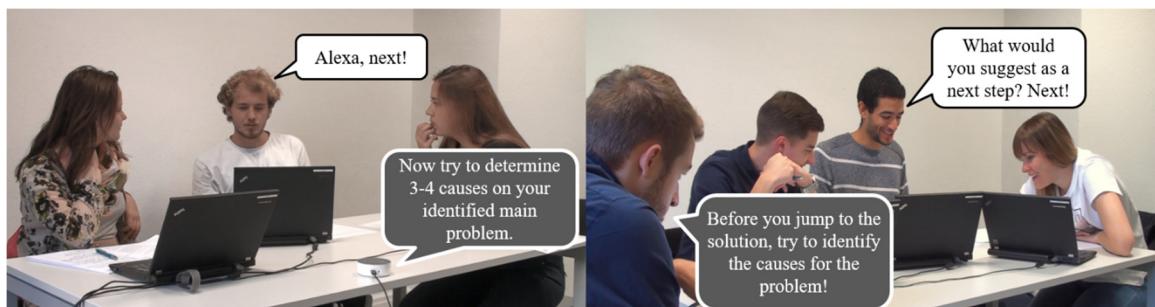


Figure 13. Example Setup with Alexa Echo Dot (left) or scripted human facilitator (right).

Source: Own illustration

The experimental procedure was divided into four components. Figure 14 outlines the procedure including activities done and the time spent on each component. In part 1, participating groups entered the room (one group at a time), received general instructions for the experiment, and were asked to complete a pretest on selected controls, such as gender, age, study background, personal innovativeness, and prior experience with SPAT.

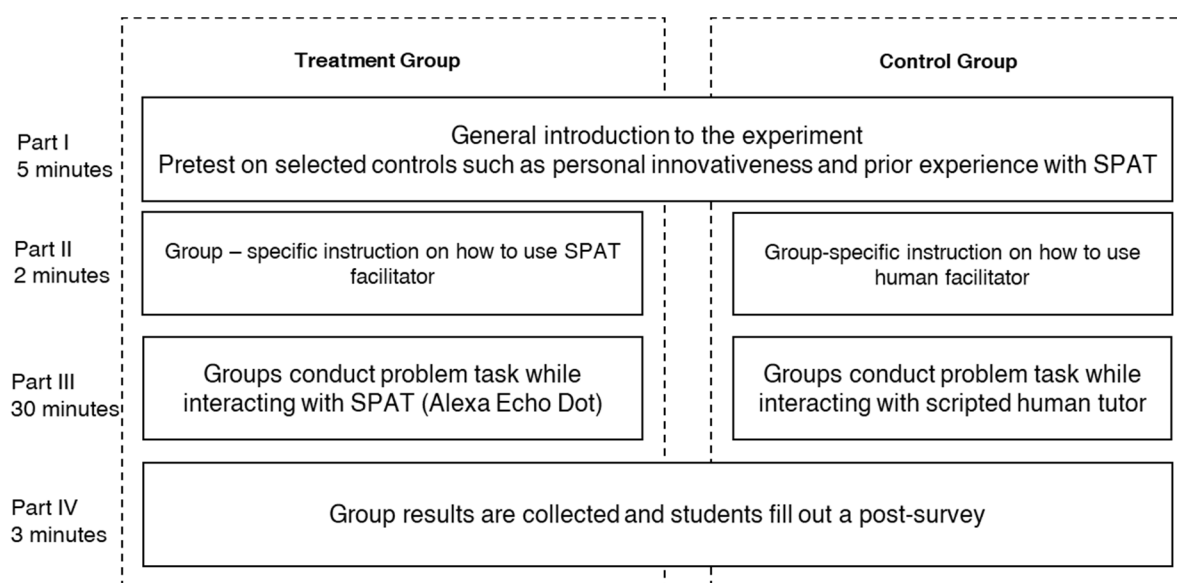


Figure 14. Experimental Procedure

Source: Own illustration

Two individuals were asked to give general instructions to the participating groups. They used a detailed script to prevent variation in the implementation. In part 2, every group received a group-specific instruction on how to use their type of facilitator. In part 3, the groups had 30 minutes to conduct the problem task while interacting with their facilitator. In part 4, I collected the group solution and the participants individually filled out a short survey containing two questions regarding the perceived helpfulness of the facilitator (from 1 to 7, manipulation check) and their overall experience of the collaboration (open question). Additionally, I recorded all the sessions on video.

5.1.4 Measurement and Analysis

5.1.4.1 Quantitative Data

For measuring the group task outcome, two experienced raters evaluated the group results using an evaluation framework aligned with the problem-solving steps of Kim and Hannafin (2011). The evaluation framework was developed collectively in a workshop within the research team. The framework contains detailed information on how well groups applied each of the five problem-solving steps, on a scale from 1 to 7, which produces the average task outcome. For example, if a group systematically derives and justifies the causes of the problem from different perspectives (problem-solving step 2: exploration), the group receives a high score near to 7. The group task outcomes were rated individually, blinded, and assessed independently of each other. Furthermore, I checked for interrater agreement with the help of a Pearson correlation. The correlation resulted in a satisfactory result (interrater agreement = 0.9699, $p < 0.05$). To measure the collaboration quality, I analyzed the cognitive and social component proposed by Graesser et al. (2018b) and Hesse et al. (2015). The cognitive component evaluates how the participants manage the task, contribute, and build knowledge

through taking appropriate action toward the solution of the problem (Hesse et al. 2015). Thus, I individually analyzed all 40 video recordings regarding (1) the frequency of applied turns and (2) the homogeneous content-specific contribution proposed by Meier et al. (2007) and Setlock et al. (2004). The frequency of turns indicates how intensively participants interact with each other. Higher turn-taking indicates that participants act more responsively to each other's utterances—that is, they reflect more on what was stated as well as elaborating or building on arguments (Drew 2005). The equality of contribution indicates the balance of the contributions of the individual group members measured by the average length of words per utterance and the corresponding SD (Hutchby and Wooffitt 2008). Hence, it is an indicator of whether the solution was developed through the joint effort of all group members or rather on the basis of a single strong participant. The two dimensions and the four components lead to a final score for the collaboration quality, as depicted in Figure 15.

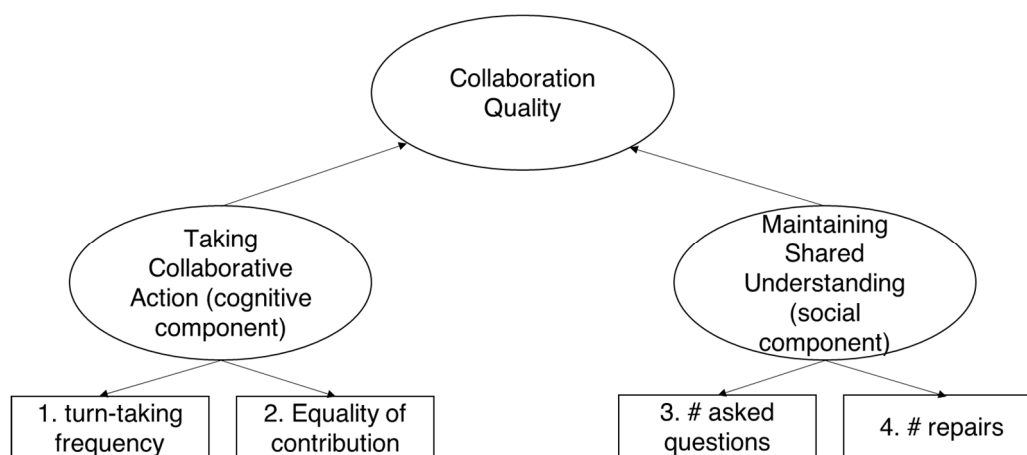


Figure 15. 4 Components Affecting the Collaboration Quality

Source: Own illustration

The social component evaluates the group members' efforts in building and maintaining a shared understanding, which is a prerequisite for effective collaboration. I measured the building and maintaining of a shared understanding with the number of on-task related questions and the number of used repairs proposed by Rummel and Spada (2005). Asking questions prevents misunderstandings and fosters a shared understanding of participants' utterances. This component was measured by the sum of question-answer pairs that occurred per session. The last measurement counts the total number of repairs per group throughout the discourse. Repairs occur whenever one participant detects a mismatch of intentions between the speaker and the recipient and subsequently corrects it (Dillenbourg et al. 2016). Hence, the participants first need to detect a misunderstanding—that is, to identify a mismatch between what one speaker intended to say and what was understood by another—and try to repair it with a question (Corti and Gillespie 2016; Dillenbourg et al. 2016). Table 23 gives an overview of all

measurement components of collaboration quality and describes how I measured them.

Measurement	Description
1. Turn-taking frequency	Average number of switches between utterances of group members
2. Equality of contribution	Average number of contributions per participant measured in average words per utterance and SD within the whole task
3. # asked questions	Average number of answered questions within the whole task incl. facilitator
4. # repairs	Average number of successfully detected and repaired misunderstandings within the whole task incl. facilitator

Table 23. Measurement of Collaboration Quality

Source: Own illustration

The collaboration quality corresponds to the equally weighted and standardized averages of these two dimensions and four components. For pre-experience with SPAT, I asked the participants how often they use SPAT on a scale from 1 to 7 where 1 is “never used” and 7 is “use it every day”. For personal innovativeness, I used the four items from van Raaij and Schepers (2008). For quantitative data analysis, I used a one-way MANOVA test to identify whether there were differences between the groups in terms of task group outcome and collaboration quality. A MANOVA assists in checking whether there is a relationship between task group outcome and the collaboration quality that might influence the hypotheses to be tested (French et al. 2008). Before using a MANOVA, I checked whether the data met the assumptions (test for normality, test for linearity and homogeneity of variance) for conducting a MANOVA. Furthermore, I calculated partial eta squared for measuring the effect size (0.01 = small, 0.06 = medium, 0.14 = large, (Cohen 1988)). I also included a post-experiment manipulation check to see whether facilitators have an overall effect on the groups by letting participants rate the perceived helpfulness of the facilitator (5.83 out of 7, no significant differences between the groups). Moreover, I used the statistic program R as a tool for analysis (Team 2013).

5.1.4.2 Qualitative Data

To gain deeper insights into how groups interact with SPAT compared to human facilitators, I used a CA approach. CA aims to uncover hidden interaction patterns

and/or give guidelines on how to improve interaction (Hoey and Kendrick 2017). The use of CA as a tool for unpacking the sequential and embedded practice of situated actions is an established technique in human-human interaction (Mackey and Gass 2011). Since computers become more interactional, conversation analysis as a method has become more popular in HCI research (Porcheron et al. 2018). To capture interaction fragments between the group members and their facilitators, I recorded all the sessions on video. The results consist of approximately 20 hours of recorded data. I focused the analysis on the groups' help requests (help function) directed to the facilitator, since this was the main reason why participants interacted with the facilitator. Overall, I identified over 175 distinct "help request" utterances—that is, talk directed at the facilitator in a seeming attempt to receive help. These requests often formed part of a larger sequence encompassing more than one help request. Thus, all in all, the corpus consists of 122 of these interaction fragments.

The collecting and analyzing of the data follow the CA approach of Hoey and Kendrick (2017) described in Subsection 3.4.3 of this dissertation. In step 1 (identify a candidate phenomenon), I analyzed how participants approach and react to the facilitators' utterances, scanning the video recordings for help requests from group members directed at either Alexa or the scripted human facilitator. In step 2 (build a collection of cases), I collected several fragments including multiple turns prior to and after the help request and followed a set of standard transcription conventions according to Jefferson's (1984) Transcription Notation. These steps show how Jefferson Transcription Notation works: I note where pauses take place ((1.4) for pauses of 1.4 seconds and (.) for pauses between words shorter than 0.2 seconds), where talk is LOUD or °quiet°, where it is spoken >faster< or <slower> than usual, and where an utterance is elongated. A period "." and a comma "," indicate a falling or slightly rising intonation contour, respectively. Overlapping talk is represented using indentation and [square brackets]; ((unspoken actions)) are noted in double parentheses; and (unclear words or utterances) are noted in single brackets. Utterances in bold indicate ones directed to the facilitator and all used names are replacements for the real names of participants to ensure anonymity. Talk is transcribed as it was produced, which means that the fragments are presented and analyzed in their original language (Swiss German or German). The results section therefore contains interaction fragments that are in the original language and translated into English for better comprehensibility. In step 3 (analyze each case in the collection), I analyzed each case individually starting with the clearest cases that show differences between the treatment and control group. In step 4 (analyze variation in the collection), I tracked forms of variation across the two TGs and sorted the cases accordingly. In step 5 (define the boundaries of the phenomenon), I looked for counterexamples to analyze the boundaries of the identified interaction patterns. In step 6 (analyze deviant cases and look for normative evidence), the goal was to describe the interaction patterns that groups use when talking to human-like SPAT compared to those

they use when talking with humans. Step 7 (produce a formal account of the phenomena) involved describing the interaction patterns in detail and explaining how they operate.

The interaction fragments I demonstrate in this paper are taken from different sub-groups in the SPAT and human facilitator groups. I present the fragments as exemplars of identified patterns. This means that they are not the sole examples of such reactions in the corpus—rather, they demonstrate patterns that have been repeatedly found in the data.

5.1.5 Results

5.1.5.1 Quantitative Results

Table 4 presents a summary of means and SDs for group task performance and collaboration quality including its individual components. Groups interacting with human-like SPATs showed both a higher collaboration quality and a higher task outcome.

	Treatment Group (SPA, n=20)		Control Group (scripted human, n=20)	
	M	SD	M	SD
1. Task Outcome	3.32	0.74	2.71	0.54
2. Collaboration Quality	5.25	1.06	4.30	1.29
2.1 Turn – taking frequency	220.83	68.32	171.80	59.20
2.2 Equality of contribution	20.97	9.66	17.44	11.49
2.3 # asked questions	55.83	21.11	33.0	8.28
2.4 # repairs	1.00	1.10	1.40	2.61

Table 24. Descriptive Statistics of Treatment and Control Group

Source: Own illustration

Before using a MANOVA, I tested assumptions. First, I conducted a test for normality and linearity as well as a test for homogeneity of variance. All assumptions were met. Next, I conducted the MANOVA test to see if group differences were significant. The results revealed that groups interacting with SPAT are able to significantly increase both group task outcomes and collaboration quality ($F(1,38)$

= 6.485, $p = 0.0040$, $\eta^2 = 0.189$, $n = 40$). Looking at each DV individually, it is apparent that groups interacting with SPAT had significantly better task outcomes ($F(1,38) = 8.444$, $p = 0.0062$, $n=40$). Moreover, groups interacting with SPAT showed significantly higher levels of collaboration quality ($F(1,38) = 6.880$, $p = 0.0292$, $n = 40$). Since these results are significant and the effect size can be considered as high, H1 and H2 are both supported (Cohen 1988).

5.1.5.2 Qualitative Results

In the following, I describe the identified interaction patterns and show exemplary interaction fragments. Again, the phenomena of interest are of participants' help requests directed at the human-like SPAT or human facilitator and the corresponding interactions. I identified three interaction patterns that differed between groups interacting with human-like SPAT compared to groups interacting with human facilitators: (1) more questions for hints and out of idle curiosity; (2) longer gaps between turns; (3) more laughter before and after SPA utterances.

5.1.5.3 Interaction Pattern 1: More questions for hints and out of idle curiosity

The first interaction pattern relates to the amount and kind of questions groups asked the facilitators. Generally, I identified three reasons why the participants approached their facilitator for help: (1) to get a clearer understanding of the task (i.e., if they were unsure about what to do and what depth of response was asked for); (2) to get an additional hint (e.g., if they were required to name three causes for traffic congestion and they could only think of two); and (3) from idle curiosity about what input a facilitator might provide. I observed that groups interacting with a SPAT facilitator requested hints and questions for reasons of curiosity more often than those working with a human facilitator. Fragment 1 in Figure 16 shows an interaction where participants asked out of idle curiosity.

...				
01→Valeria:	→	Sollen wir aus Interesse Alexa fragen?	(1.4)	Should we ask Alexa out of curiosity?
02→	→			
03→Nicolas:	→	Ich...		I
04→Valeria:	→	Weil ich also		Because I
05→Nicolas:	→	bin dafür. (.) Alexa mal auszuprobieren		Agree to try Alexa
06→	→	[[(laucht)]]		[[(laughing)]]
07→Valeria:	→	[[Ja...]] weil also brauchen die		[Yes] because need, well the tasks are
08→	→	Aufgaben sind einfach, brauchen tut		straightforward, so one does not really
09→	→	man die eigentlich nicht, (oder)?		need her, (right)?
10→Nicolas:	→	Ja genau		Yes exactly
11→	→	(1.4)		
12→Valeria:	→	hmm		Hmm
13→	→	(1.5)		
14→Nicolas:	→	Also Alexa (1.6) <u>hilfe</u>		Alright Alexa <u>hilfe</u>
15→	→	(2.0)		
16→Nicolas:	→	oh ok. → [[(laucht)]]		Oh ok. [[(laughing)]]
17→Anne:	→	→ [[(laucht)]]		[[(laughing)]]
18→Valeria:	→	→ [[(laucht)]]		[[(laughing)]]
19→Nicolas:	→	Alexa (2.1) <u>help</u>		Alexa <u>help</u>
...				

Left = Original Excerpt in German

Right = English Translation

Figure 16. Fragment 1: Asking for Idle Curiosity, TG

Source: Own illustration

5.1.5.4 Interaction Pattern 2: Longer Gaps between Turns

The second pattern I identified relates to the gaps after the facilitators' utterances. As can be noted from interaction Fragment 2, interactions between participants and SPAT (TG) involved a mutual silence period of two seconds or longer (see line 12, Fragment 2) after the facilitator's input, before they jointly discussed the input received. A contrary situation was observed in groups interacting with a human facilitator as interaction Fragment 3, line 15, shows. The difference in the length of produced silence between the two TGs holds across nearly all sessions.

...	¶		
01 → Sandro:	→	Wollen wir sie fragen?	Do we want to ask her?
02 →	→	((zeigt auf Echo))	((points at the Echo device))
03 → Jonas:	→	Ja, tuen es mal probiere.	Yes, try it.
04 →	→	(0.8)	¶
05 → Sandro:	→	Alexa, Help. Was ist eine Lösung?	Alexa, Help. What is a solution?
06 →	→	(1.8)	¶
07 → Alexa:	→	Okay, ich gebe euch einen Hinweis.	Okay, I will give you a hint. A possible
08 →	→	Eine mögliche Lösungsoption könnte im	solution might relate to traffic
09 →	→	Bereich Verkehrsregulierung sein.	regulation, as for example roundabouts.
10 →	→	also zum Beispiel Kreisverkehre. Was	What could be further solutions?
11 →	→	könnten weitere Lösungsoptionen sein.	¶
12 →	→	(2.2)	Well, an intersection
13 → Jonas:	→	Joh: (2.3) (e-chrüzig)	
...	¶		

Figure 17. Fragment 2: Breaks after SPAT Utterance, TG

Source: Own illustration

...	¶		
01 → Johanna:	→	Fragen wir nach Hilfe?	Let's ask for help?
02 → Elisa:	→	Ja	Yes
03 → Nicole:	→	Ok. Facilitator, help us. Wie finde	Ok. Facilitator, help us. How to find the
04 →	→	ich die beste Lösung?	best solution?
05 →	→	(0.9)	¶
06 → Facilitator:	→	Wenn ihr das Gefühl habt, dass die	Okay, I will give you a hint. If you
07 →	→	Lösungsoption das Kriterium gut	believe, a solution fulfills the
08 →	→	erfüllt, dann setzt eine Vier oder	criterion well so give it a five and
09 →	→	Fünf und sonst eine andere Zahl.	otherwise a one. If we take roundabouts
10 →	→	Wenn wir als Beispiel Verkehrs-	as an example, then costs are within
11 →	→	regulierung und da Kreisverkehre	reasonable limits and would account for a
12 →	→	nehmen, dann halten sich die	four.
13 →	→	Kosten als Kriterium im Rahmen und	¶
14 →	→	würde zum Beispiel eine Vier ergeben.	¶
15 →	→	(0.6)	¶
16 → Johanna:	→	Ah, Kosten sind ein Kriterium zum	Ah, costs are a criterion for example
17 →	→	Beispiel	¶
18 → Elisa:	→	Ah	Ah
19 →	→	(0.6)	¶
...	¶		

Figure 18. Fragment 3: Breaks after Human Facilitator Utterance, Control Group

Source: Own illustration

5.1.5.5 Interaction Pattern 3: More Laughter Before and After SPA Utterances

The third pattern I identified was that participants laughed more before and after help requests directed to the SPAT, as can be seen from Fragment 4, lines 13–16. In the human–SPAT group, the participants almost never responded with laughter to the humans' advice.

...			
01→Marc:	→	Alexa stop.	Alexa stop.
02→	→	(2.4)	
03→Alexa:	→	Wie kann ich euch helfen? Sagt next oder	How can I help you? Say next help or
04→	→	help oder repeat.	repeat.
05→Marc:	→	Help.	Help.
06→	→	(1.4)	
07→Alexa:	→	Okay, ich gebe euch einen Hinweis.	Okay, I will give you a hint. A potential
08→	→	Eine mögliche Ursache ist zum Beispiel	cause is for example the lack of ring
09→	→	steigendes Verkehrsaufkommen. Was	roads or public transportation. What
10→	→	könnten weitere mögliche Ursachen	could be further causes.
11→	→	sein.	
12→	→	(1.0)	

			[[(laughing)]]
			[[(laughing)]]
			[[(laughing)]]

Figure 19. Fragment 4: Laughing after SPA Utterance, TG

Source: Own illustration

5.1.6 Discussion

Using a mixed-method approach, this study evaluates the use of the DP *human-ness* incorporated in SPAT as facilitators in a CPS environment. The quantitative findings suggest that human-like SPAT has the potential to outperform human facilitators when both have the same scope of information and functions. The qualitative findings provide more in-depth insights into possible reasons for this. Specifically, the quantitative analysis shows that groups interacting with SPAT facilitators took turns more frequently, indicating that these groups had a more vivid discussion and made a more intense contribution than those interacting with a human facilitator. One reason for this might be that participants feel more relaxed and open when there is no human facilitator to judge them (Pickard et al. 2016). This concurs with my identified interaction pattern 1, which shows that groups interacting with SPAT asked more questions to receive hints and out of idle curiosity. Consequently, they showed a higher willingness to include the SPAT facilitator in their collaborative activities.

Furthermore, the SPAT groups made a more equal contribution. For example, a sole participant seldom addressed the SPAT facilitator without the consent of at least one other group member. The regulation of whether to initiate the interaction with the SPAT, and who should do the initiating, is therefore already a collaborative achievement and prompts the contribution of several participants. One reason for a more equal contribution might be that all participants feel more confident and comfortable interacting with SPAT facilitators than they do when interacting with human facilitators. Similar effects can be observed in the field of language learning, where students feel more open talking to a machine than they do when talking to a human facilitator (Wik and Hjalmarsson 2009). It seems that group members feel a higher degree of agency when getting support from a computer rather than a human facilitator leading to a more efficient collaboration (Limerick et al. 2014). This is also reflected by my identified patterns 1 and 3, which indicate that participants asked more questions and laughed more during interaction with the SPAT facilitator, suggesting that they are more open and unrestricted in these situations.

Moreover, the quantitative data revealed that groups interacting with SPAT facilitators put greater effort into the achievement and maintenance of a shared understanding of the problem between group members. This is indicated in the total number of asked and appropriately answered questions about content and about approval for their arguments. These results are confirmed by the conclusions of Roschelle and Teasley (1995), who stated that successful collaboration involves a large degree of mutual engagement, joint decision making, and discussion. Conversely, the quantitative results also showed that groups interacting with a human facilitator experienced more repaired misunderstandings. This concurs with the findings of Corti and Gillespie (2016), who noted that participants put less intersubjective effort (assessed on repair activity) into interacting with an artificial agent than with a human being. Moreover, the qualitative data revealed that participants experienced longer pauses after receiving advice from the SPAT. After more than two seconds, they reacted to the advice and continued their discussion, whereas groups receiving advice from the human tutor reacted immediately. One reason for these produced silences might be that group members listen to the SPAT facilitator more carefully, whereas advice from a human facilitator is directly acted upon. This might arise from politeness, as longer pauses in human-to-human conversation are perceived as negative (Porcheron et al. 2017). Another explanation might be that participants feel less pressure from the SPAT than they do with a human facilitator and therefore take time to reflect on what has been said. Our findings agree with Sack et al.'s (1978) substantial auditory analyses of conversational data, revealing that human-human communication typically occurs without any silence between interactions and without any overlapping speech ("no-gap-no-overlap"). One other explanation for the differences might be that participants are better able to detect turn-ends made by a human compared to those made by a SPAT facilitator and can therefore prepare their reactions faster (Heldner and Edlund 2010).

To summarize, these findings suggest that humanness incorporated in SPAT enables these systems to approach the role of a human tutor. The qualitative data further suggests that SPAT might have some advantages over human tutors (at least for scripted human tutors). My work makes several theoretical contributions. First, it contributes to computer-supported collaboration by showing that new emerging SPAT can be successfully used for supporting CPS environments. Although I did not use a freely interacting facilitator as a baseline, the findings indicate that SPAT might be able to directly compete with human facilitators in the near future. Given that the technology around SPAT is developing rapidly, it can be considered as a useful alternative to cost-intensive human facilitators. Moreover, my work also suggests some of the benefits that SPAT facilitators may have compared to human facilitators. The findings can thus be considered as a starting point in helping to determine the possible benefits of SPAT facilitators compared to human facilitators. For example, group members appear to be more open and feel higher degrees of agency when interacting with SPA facilitators. These insights provide the basis for future hypothesis development in which researchers

can investigate certain dimensions of humanness that reveal how SPAT characteristics facilitate group collaboration more effectively than human facilitators. As SPAT increasingly enters our everyday lives, it is important to understand how we can leverage its potential beyond the existing, traditional use cases. Second, by applying a CA approach, I have shown that there are hidden interaction patterns when interacting with SPAT rather than humans. These findings demonstrate that conversation analysis should be used more often in IS research to uncover hidden interaction patterns that are not visible with quantitative or frequently used qualitative methods (e.g., interviews). The insights resulting from this method have the potential to inform future research in the area of HCI. Regarding the practical implications of this study, I have suggested how SPAT designers might integrate humanness DPs into SPAT for collaborative problem solving beyond already known use cases. It is worth focusing on use cases of this kind, since collaboratively solving problems will become increasingly important in future job environments, and SPAT is becoming more ubiquitous. Finally, I have given SPAT designers some first impressions of the differences between human–SPAT and human–human communication, which will help to inform future designs of SPAT.

5.1.7 Limitations and Future Research

My study has several limitations. The primary limitation relates to the control group. I compared the value of SPAT with scripted human facilitators. It is to be expected that freely interacting human facilitators would have led to different results. However, I tried to ensure high quality support by aligning the used script to the well-known problem-solving steps described by Kim and Hannafin (2011) and also included a manipulation check indicating that participants perceived the support as helpful. SPAT technology is not yet capable of building a free interaction with the user. Nevertheless, rapid developments in machine learning and NLP are promising and indicate that SPAT will soon be able to function as a self-sufficient collaboration partner or facilitator (Seeber et al. 2018). The second limitation concerns the small sample size of 40 groups. However, by using a mixed-method approach, I was able to partially compensate for that (Venkatesh et al. 2013). Despite the small sample size, the statistical tests were able to reveal significant differences between groups interacting with SPAT and those interacting with human facilitators. It would be interesting to see if further research involving larger sample sizes can confirm these results. A third limitation is that there might be within-group dependencies that increase the likelihood that the analysis will incorrectly suggest a significant effect (Baldwin et al. 2008). However, the goal of this study is to show that SPAT has an effect on social interaction. Hence, I accounted for intra-group dependencies with collaboration quality as one of my DVs. A fourth limitation is that some interaction with the SPAT is driven by curiosity. This means that I cannot be fully confident that the effects are equal when curiosity levels are the same in both groups. However, I asked participants about their pre-experience with SPAT, which resulted in a rather high experience level. A fifth limitation is that I investigated the effects in a narrow computer-supported

collaboration context (face-to-face, CPS). Further research should try to replicate the results with other kinds of collaboration settings (other tasks, group sizes, etc.). A sixth limitation is that the research design and analysis does not allow a full investigation of why human-like SPAT facilitators perform better than human facilitators. Why is SPAT performing better than human facilitators? Are there differences in terms of evaluation fears or trust? Future research should consider finding answers to these questions. Finally, it would be interesting to see if SPAT can also keep pace with expert human facilitators and whether other researchers can detect different interaction patterns with the help of CA methods.

5.1.8 Conclusion and Next Steps for the Course of the Dissertation

This study focused on the value of humanness in SPAT in a CPS context. Using a mixed-method approach, I first identified relationships quantitatively and enriched the data with a conversation analysis approach. The findings suggest that human-like SPAT is able to improve participants' task performance and also their collaboration quality and is able to outperform scripted human tutors. Moreover, I was able to uncover some specific interaction patterns that differ between human–SPA and human–human communication. The findings contribute to developments in computer-supported collaboration and emphasize the use of conversation analysis as a method of uncovering often hidden interaction patterns in HCI.

This study provides evidence that humanness in SPAT can help students to improve outcomes in a laboratory and a collaborative learning setting. Building on this, I incorporate humanness in the studies described in the following chapters and test other relevant DPs that I derived in the course of the study in section 4.2. The next study completes RQ2a and focuses on conversational interactivity as DP, switching from an artificial environment to a real blended learning environment.

5.2 Using Smart Personal Assistant Technology to increase Long-term Skill Development – Insights from a High School and Vocational School

5.2.1 Introduction

This study will complement RQ2a and focuses on *conversational instruction* (DP2) in SPATs. The study takes place in a real blended learning environment and focuses on skill development in a 10th grade high school and vocational business school. Thus, it also helps to answer RQ2a:

How can Smart Personal Assistant Technology improve students learning success in blended learning environments?

The initial starting point for the answer of this RQ are the formulated DPs in section 4.2 and especially the DP 2, conversational instruction. Conversational instruction is the primary way to change text to emphasize first- or second-person forms of address (e.g., “you”, and “your”) rather than third person forms (e.g., “the”). Moreover, a conversational instruction is described by multiple turns within a conversation between the student and the educator trying to build on students’ pre-knowledge and pre-experience. This is in accordance with the constructivist learning paradigm described in section 2.2. As both of the above-mentioned aspects relate the instructional material to the student, an individual who learns from it is more likely to become interested and engaged in the mental processing of the material (the self-reference effect, Klein and Loftus 1988), thus resulting in improved learning. I translated the mentioned *conversational instruction* DP into DFs and integrated it into a SPAT.

In summary, the majority of existing empirical research in the area of SPAT and education is rather explorative (e.g., observations, qualitative interviews with students and educators) and few empirical studies that rigorously measure the effectiveness of SPAT in learning environments exist (Dousay and Hall 2018; Jean-Charles 2018). In particular, empirical studies are lacking when it comes to longitudinal effects, e.g., whether SPAT can help to train certain skills over the course of a longer period of time. This is crucial, since we know from prior research that the development of skills take time and is difficult to achieve within a single short experiment (Soderstrom and Bjork 2015).

5.2.2 Method

To answer the research question, I create a quasi-experimental design, where I complement the quantitative results of two field quasi-experiments with the qualitative insights of a post-experiment focus-group discussion. Quasi-experimental designs in education are adopted in cases where random formation of groups is not possible and therefore experimental and control groups are formed with already-existing classes (Fraenkel et al. 2011). In two different schools, I conducted pre-tests to check whether classes within each school are similar in terms of initial

problem-solving skills, gender, age, school grades in the relevant subject, personal innovativeness, and pre-experience with SPAT. The pre-test scores showed that within each school, the two classes were similar. Thus, I was able to randomly choose one to be the experimental class that follows an intervention program in each school. Detailed pre-test results and student characteristics can be seen in Appendix A.3 and A.4.

5.2.2.1 Background and Setting

I selected classes from a second-year high school (October/November 2018) and a second-year vocational business school (January/February 2019) for the experimental and control groups. The choice of implementing SPAT in these learning environments has two main benefits. First, in each school, the experiment class and control class were similar. Moreover, in each school, the two classes had the same teacher and the same learning content, making the two classes suitable for a field quasi-experiment approach. Second, choosing two different school types allowed me to widen the scope of our study and strengthen our results. In a high school, school performances are usually better as students aim to prepare themselves for university. In a vocational business school, school performances are generally a bit lower as students tend to start to work after school. Both schools I chose are located in the capital city of its region. The high school had approximately 1,300 students in total and eleven second-year classes and the vocational business school had 500 students in total and five second-year classes. Both experiment and control classes took courses on the “Introduction into law”, which is a unit in the business and law curriculum. I chose this unit for two reasons. First of all, it is the first occasion that students come into contact with law, therefore students have no pre-knowledge and the classes are comparable. Second, the unit includes the solving of legal problems and is therefore very suitable for the acquisition of problem-solving skills (the focus of the study).

The experimental groups completed the module with SPAT, while the control groups completed the same module with traditional paper-based learning materials. Within each school, the same teacher taught the experiment class and the control class. The relevant teacher in both schools was very experienced, having taught for over 25 years in different types of schools (elementary school, high school, vocational school, etc.). They reported using technology with their students daily (computer-based teaching, tablet-based teaching, etc.). Over the experiment period of five weeks, both classes in both schools had to complete the same module with the same six learning goals (LG). These learning goals are depicted in Appendix A.1. The teachers used the same teaching methods in both classes, consisting of frontal teaching, followed by individual and partner work. None of the students in the four classes had had law-related subjects before.

5.2.2.2 Sample

Following Sahin and Yilmaz (2020), I used convenience sampling in this study, meaning that I chose the schools from two different school types that had two comparable classes and were willing to participate in the study.

Experiment 1 in High School

The sample in experiment 1 consisted of students in two second-year classes of a high school. The experiment class had 9 males and 13 females, with an average age of 16.9 years. The control class had 11 males and 12 females, with an average age of 17.1 years. I conducted ANOVA tests to make sure that the experiment and control class were similar. The tests revealed no significant differences regarding gender ($p = 0.650$), age ($p = 0.788$), school grades in business and law ($p = 0.558$), pre-experience with SPAs ($p = 0.941$), personal innovativeness ($p = 0.191$), and the pre-test problem-solving skill results ($p = 0.774$). This experiment class did not participate in the post-experiment focus group discussion.

Experiment 2 in Vocational Business School

The sample in experiment 2 consisted of students in two second-year classes of a vocational business school. The experiment class had 12 males and 10 females, with an average age of 17.4 years. The control class had 12 males and 11 females, with an average age of 17.2 years. I conducted ANOVA tests to make sure that the two classes were similar. The tests revealed no significant differences in the background of the students in the two classes in terms of their gender ($p = 0.248$), age ($p = 0.743$), school grades ($p = 0.100$), pre-experience with SPAT ($p = 0.552$), personal innovativeness ($p = 0.332$), and the pre-test problem-solving skill results ($p = 0.376$). All students of the experiment class of this school participated in the post-experiment focus group discussion.

5.2.2.3 Task and Smart Personal Assistant Design

Between the end of week 1 and the end of week 4, the students had to do four 30-minute, problem-based, homework assignments for learning goals (LGs) 2, 3, 4 and 5. LG 1 was part of the in-class introduction of the related subject. The homework assignments were designed and developed together with the two class teachers. The homework assignments contained problem tasks related to the difference between morality, custom, and law as well as the fundamental rights “freedom of opinion”, “freedom of religion”, and “property guarantee”. All four classes received the task catalog as a paper-based script. The students from the experiment classes used Amazon’s Alexa for the homework assignments. The students from the control classes received an additional paper-based script with steps and hints that are also included in the Alexa Skill. All the homework assignments were in the same style as the pre- and post-test tasks. Furthermore, they fulfilled the requirements of a problem-based task according to Jonassen.

Accordingly, the tasks were open-ended, unstructured, resonated with the students' experience, and realistic. As an example, in Appendix B, we show homework assignment 2 and how it addresses these requirements.

The main goal of this experiment was to evaluate DP2, conversational instruction. To incorporate conversational instruction into the developed SPAT, we used the AVS and the underlying Amazon Web Services (AWS) for two reasons. First of all, AVS offers an intuitive graphical interface to design the SPAs (Amazon 2019). Second, AVS seems to offer one of the most developed State-of-the-Art capabilities regarding natural language understanding and processing. The database (data related to the problem-solving steps, Alexa's answers for each problem-solving step with each assignment, etc.) and corresponding functions are developed and stored in AWS Lambda.

To implement DP2, the dialog flow uses a proactive and reactive logic to make sure that multiple turns occur and the student also asks individual questions. In the proactive logic, the SPAT provided instructions that guided students through the task by using five different problem-solving activities adapted from Kim and Hannafin's (2011) problem-solving phases (see grey rectangles on the left side in Figure 20): *(a) problem identification and engagement*, *(b) problem exploration*, *(c) problem reconstruction*, *(d) solution presentation and communication*, and *(e) reflection and negotiation*. For example, in a first step (related to problem identification and engagement), the SPA asked the students to identify the core of the problem. These instructions were the same for each student and each task. Alexa has started with the first step as soon as the students have chosen one of the four homework assignments.

In the reactive logic, the SPAT answered students' questions while working on the problem steps. With the help of AWS Lambda functions, I tracked the current progress of the students. The SPAT was therefore able to recognize whether an answer was wrong or right and responded accordingly. For example, in Figure 20, the SPAT gave a "Are you sure ..." response. Moreover, the SPAT was able to provide feedback to three different types of questions or queries. Type 1 was the students' query to skip to the next step. Type 2 was content-specific questions (concept clarifications, asking for hints, etc.). In Type 3, the students requested to repeat the current problem-solving step (see grey rectangles on the left side). For example, in Figure 1, the student asked a content-specific question whether it is public law and the SPAT provided feedback. Whenever Alexa was not able to understand the students, it invoked a fallback intent that allowed students to rephrase the question or jump to the next problem-solving step.

The proactive and reactive logic helps to imitate a human-led conversational instruction, thereby internalizing the problem-solving steps, which should result in a gain of their problem-solving skills. As an example, Figure 20 shows an excerpt of a student dialogue and the corresponding proactive and reactive functions.

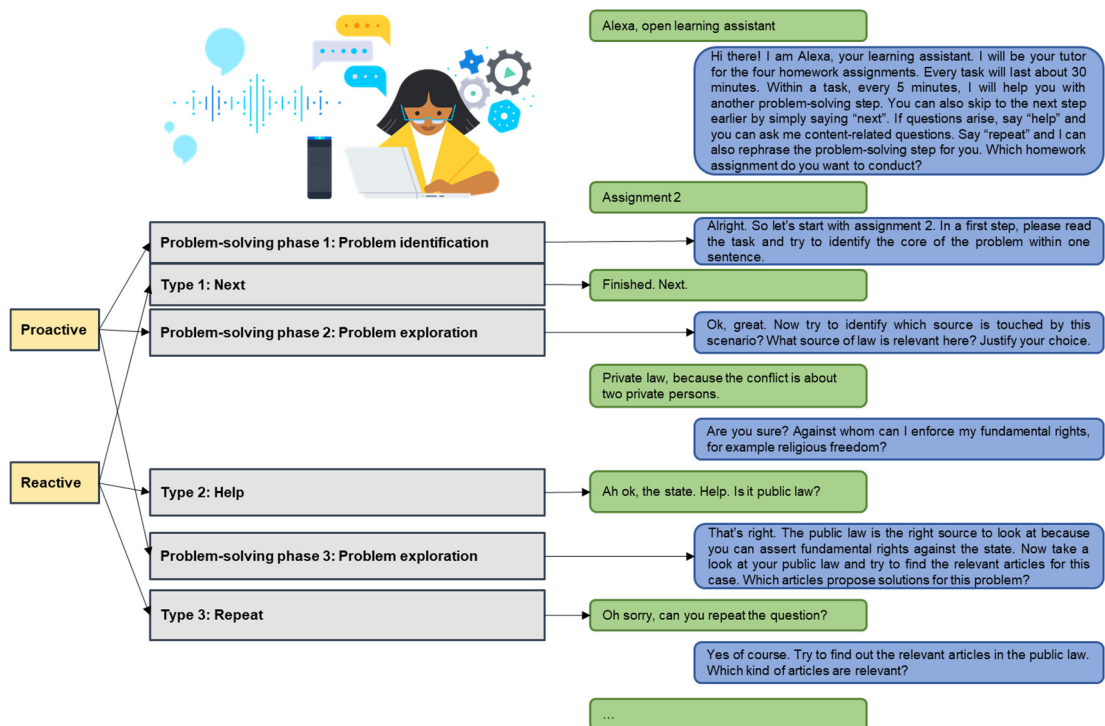


Figure 20. Example dialogue between the Student and SPAT, and Corresponding Proactive and Reactive logic.

Source: Own illustration

A key challenge when designing SPAT-based learning materials for experimental treatments is information equivalence. Larkind and Simon (1987) argue that two representations are informationally equivalent if all the information from one representation can also be inferred from the other representation and vice versa. This ensures that differences stem from the technology itself, rather than the content of the treatment. Therefore, we included all the information incorporated in the SPAT also on the paper-based learning materials provided for the control class. Students in the control class received an additional script (separate from the task script) containing all the steps and hints that are also included in the Alexa Skill. Figure 21 shows parts of the additional script, which is the same excerpt as in Figure 20.

Assignment 2:

Step 1: Let's start with assignment 2. In a first step, please read the task and try to identify the core of the problem within one sentence.

Step 2: Try to identify which source is touched by this scenario? What source of law is relevant here? Justify your choice.

Step 3: Take a look at your public law and try to find the relevant articles for this case. Which articles propose solutions for this problem?

...

Hints:

....

Step 3: The public law is the right source to look at because you can assert fundamental rights against the state.

Figure 21. Excerpt of additional Script for Homework Assignment 2 provided for Control Group.

Source: Own illustration

5.2.3 Experimental Procedure

Figure 22 depicts the study timeline for experiments 1 and 2.

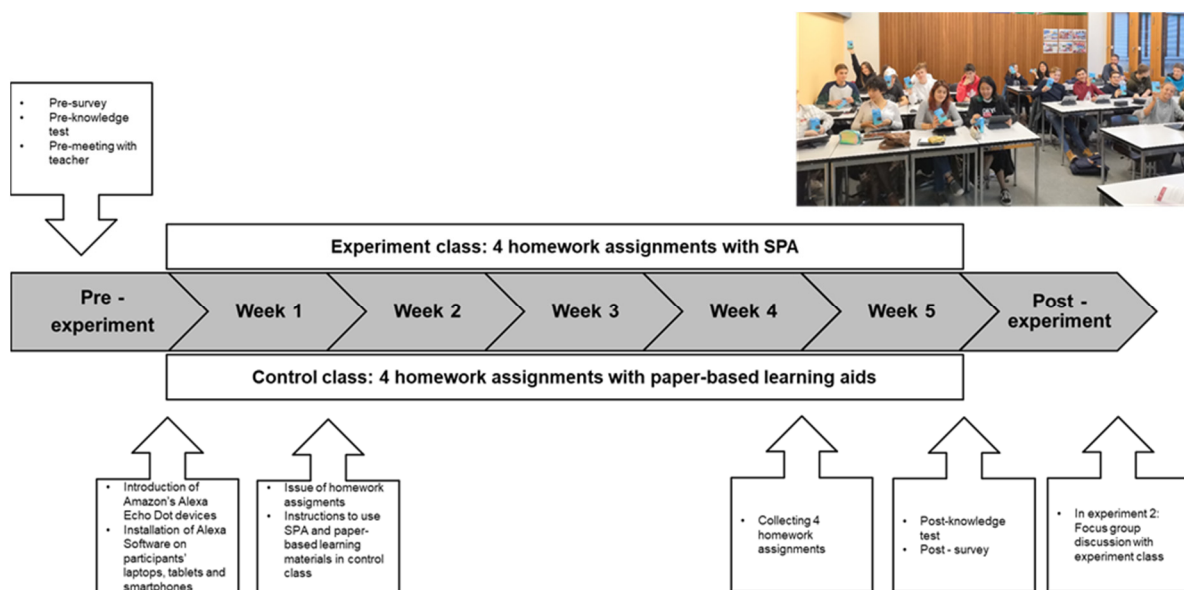


Figure 22. SPA Experiment Implementation Timeline

Source: Own illustration

Several tasks were conducted before the experiment started, i.e., during the pre-experiment state. During a lesson, all students from all four classes completed a survey to help identify and control for pre-existing class differences (see Appendix C). Also, the classroom teachers administered a 30-minute, 3-subtask pre-

knowledge test to all students to compare the knowledge level of each class. Moreover, the research team organized a pre-experiment meeting with the teachers to discuss the experiment details with them and to ensure that both classes are taught the same way over the period of the experiment.

In week 1, I installed the Amazon Alexa Echo Dot devices in the experiment classes and helped students to install the accompanying Alexa software on their smartphones, tablets, and laptops. At the end of week one, I ensured that students in the experiment classes had access to Alexa on one or more devices. After week 1, all four classes received their four homework assignments in the form of a paper script (non-digital). The experiment classes were instructed to use Alexa on their preferred devices (standalone device, smartphone, tablet and laptop) while conducting their homework assignments. The experiment group students did not use SPAT during class. The control classes were instructed to use the paper-based learning materials that were sent to them in form of an additional script (with exactly the same amount of information as on Alexa). At the end of week 4, all four classes submitted their homework assignments. One week later, they took their 30-minute post-knowledge test and post-experiment survey. The teacher was instructed to not discuss the homework assignments before the students conducted the post-knowledge test to ensure that the final results were not corrupted. In the post-experiment phase, the experiment class in the vocational business school participated in a 45-minute focus group discussion.

5.2.4 Measurement and Analysis

5.2.4.1 Quantitative Data

For measuring problem-solving skills (DV), I constructed a 3-task pre- and post-experiment tests. The tasks were similar to the homework assignments, all addressing the requirements of problem-based tasks proposed by Jonassen (2000). The pre-test and post-test had the same number of tasks and possible top score and the tasks addressed learning goals 2, 3, 4 and 5 (see Appendix A.5). All tasks require the application of Kim and Hannafin's (2011) problem-solving steps. Two teachers compared the pre-test and the post-test in terms of difficulty and found them to be similar. I analyzed pre-test and post-test results with three experienced raters independently and blinded with a pre-defined and commonly-discussed rating framework. One of the raters was the class teacher. The rating framework allowed me to evaluate the application of the problem-solving steps proposed by Kim and Hannafin (2004), giving points to each well-applied problem-solving step within a task (see Appendix A.6). The final scores of the pre- and post-test results arose from an average of the individual appraisals. The maximum number of points to be achieved was 36: i.e., 12 points per task. Moreover, I calculated gain scores as the difference between a students' posttest and pre-test scores. I checked for inter-rater agreement on the pre- and posttest scores of the three raters with the help of Light's Kappa for multiple raters for both classes in both schools. The values for Study 1 and 2 are depicted in Table 25. The strength of agreement refers to the recommendation of Cohen (1960).

	Study 1: High school			Study 2: Vocational school		
	Kappa	P-value	Strength of Agreement	Kappa	P-value	Strength of Agreement
Pre-knowledge test	0.843	< 0.01	Very Good	0.742	< 0.01	Good
Post-knowledge test	0.784	< 0.01	Good	0.610	< 0.01	Good

Table 25. Interrater - Agreement Scores

Source: Own illustration

The pre-experiment survey consisted of items regarding the students' pre-experience with SPA, personal innovativeness, gender, and age. I asked students how often they use SPAs (e.g., Apple's Siri) in a week to measure their pre-experience with SPAs. For personal innovativeness, I used the four items from van Raaij and Schepers' (2008) scale. The detailed pretest-results and the detailed information about the similarity of the two classes can be found in Appendix A.8. The post-experiment survey included two open-ended questions. Question 1 was about the students' perceived helpfulness of the learning aid (SPA or paper-based learning materials). Question 2 was treatment-specific, asking the experiment classes about their experiences with the SPAT used. The items of the post-survey are shown in Appendix A.7.

To analyze the quantitative data, I conducted a one-way ANCOVA (analysis of covariance), including the pre-test scores as a covariate. ANCOVA helps to analyze variances between the groups and control for covariates. It is therefore suitable for pre- and post-test designs (Dimitrov and Rumrill 2003). I calculated Cohen's d (1988) to show the effect size. Moreover, I calculated 95% confidence intervals and used the statistic program R as a tool for analysis (Team 2013).

5.2.4.2 Qualitative Data

The focus group discussion has built on the open questions in the post-survey and addressed RQ2. It aimed to gain a more in-depth understanding of how SPA technology affects students' learning processes. The discussion lasted 45 minutes and had one of the researchers as the facilitator. All students from the experiment class in the vocational business school participated. The focus group included the following steps. First, the facilitator introduced the goal of the group discussion (to gather students' perceptions of using SPAs while learning). Second, students were asked to divide into workgroups of four to six and were invited to discuss and negotiate opinions about how they used the SPAT during their

homework assignments. To help students structure their discussion, they received three broad subject areas identified from the post-survey responses: Pros of using SPAs, cons of using SPAs, and neutral observations while interacting with SPAT. Finally, a plenary discussion moderated by the facilitator encouraged further discussion. I recorded the session, transcribed it, and used a thematic analysis to induce topics following the method of Ryan & Bernard (2003). Specifically, I used the keywords-in-context method for this study. With the help of this technique, I identified keywords indicating some aspects of the learning process with SPAT and then systematically searched the corpus of the transcribed text to find all instances of the word or phrase. Each time I came across an instance of the word or phrase, I made a note of it and its immediate context. I identified themes by physically sorting the instances into piles of similar meaning (Ryan and Bernard 2003). Then, I conducted a respondent validation by getting participants to verify the identified themes.

5.2.5 Quantitative Results

5.2.5.1 Experiment 1 in High School

Table 26 shows the means and SDs of the pre-knowledge and post-knowledge test results and the gain scores.

	Experiment Class (SPAT)			Control Class (traditional learning materials)		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Pre-knowledge test	10.46	1.67	23	10.27	2.44	22
Post-knowledge test	23.15	4.33	23	20.63	5.60	22
Gain scores	12.69	3.74	23	10.36	3.51	22

Table 26. Summary of Means

Source: Own illustration

To check that the assumptions for the ANCOVA model are met, I conducted statistical and visuals test for normality, homogeneity of variance, and homogeneity of regression slopes (see more details in Appendix A.8). In particular, I conducted a Shapiro-Wilk test to test the normality assumption ($p = 0.8927$), a Levene's test to check for the homogeneity of variance assumption ($p = 0.7706$), and an interaction analysis for the homogeneity of regression slopes assumption ($p = 0.601$). The test results (all are not significant) as well as the plots indicate that the assumptions for conducting an ANCOVA are met. We ran the ANCOVA with post-test scores as the DV, the TG as the independent variable, and the pre-test score

as a covariate. The tests show that, controlling for the pre-test, there is a significant relation between SPAT usage and problem-solving skills: $F(2, 42) = 4.514$, $r^2_{adjusted} = 0.5255$, $p = 0.0396$, *confidence intervals for mean difference* = -0.5929 and 5.6329, $N = 45$. Cohen's d is 0.5178, indicating a difference between the means of 0.5178 SDs. Since this effect is considered as medium (Cohen 1988), I can assume that student interactions with SPAT have a positive effect on acquiring problem-solving skills compared to paper-based learning materials. Moreover, a comparison of the gain scores reveals that participants in the SPAT group learned significantly more than participants in the paper-based learning materials group ($p = 0.0400$).

5.2.5.2 Experiment 2 in Vocational Business School

Table 27 shows the means of the pre-knowledge test and post-knowledge test results, including the SDs and gain scores.

	Experiment Class (SPAT)			Control Class (traditional learning materials)		
	Mean	Std. Dev.	N	Mean	Std. Dev.	N
Pre-knowledge test	11.7	1.63	22	10.67	2.16	23
Post-test	27.95	3.27	22	21.88	3.67	23
Gain scores	16.25	3.50	22	11.21	3.74	23

Table 27. Summary of Means

Source: Own illustration

The statistical test results (Shapiro-Wilk test for normality assumption, $p = 0.7847$, Levene's test for homogeneity of variance assumption, $p = 0.7203$, test for homogeneity of regression slopes, $p = 0.260$) and plots given in Appendix A.8 indicate that all assumptions for conducting an ANCOVA are met. I ran the ANCOVA with the post-knowledge test scores as the DV, the TG as the independent variable, and the pre-knowledge test score as a covariate. Results of the ANCOVA indicate that, controlling for the pre-test, the relation between SPAT usage and problem-solving skills is highly significant: $F(2, 42) = 30.573$, $r^2_{adjusted} = 0.42$, $p = 1.88e-06$, *confidence interval for mean difference* = 3.9172 and 8.2228, *Cohen's d* = 1.70, $N = 45$. Since the Cohen's d for this result is considered as high (Cohen 1988), I assume that students' interactions with SPAT have a positive effect on acquiring problem-solving skills compared to the use of traditional learning materials. Moreover, a comparison of the gain scores reveals that participants from the SPAT group learned significantly more than participants from the paper-based learning materials group ($p = 8.17e-06$).

5.2.6 Qualitative Results

5.2.6.1 Focus Group Discussion

Since our quantitative findings lead me to believe that interactions with a conversationally instructive SPAT have a positive effect on problem-solving skills, I now investigate *how SPAT affect students' learning processes* to gather deeper insights into the observed effect. The three main themes I identified from the focus group data are *interaction*, *usage behavior*, and *individualization*. The main topics and subtopics and the corresponding frequencies of the statements are presented next (see Table 28) and further elaborated in the Discussion section.

Identified topics and subtopics (out of a student's perspective)	Frequency of Statements
I. Interaction	
- I felt like having an interaction with a real tutor.	22
- Receiving challenging questions and follow-up questions triggered my thinking processes.	15
- SPA responded immediately when I asked a question.	9
II. Usage Behavior	
- Very convenient to use it on my smartphone similar to Google's Assistant or Apple's Siri.	11
- I also used the SPA on other learning places where I would normally study.	7
III. Individualization	
- Felt like the SPA nearly always detected what my questions were about.	13
- It was great that the SPA remembered the current status when I continued to work on a task.	7
- SPA only helped when I needed support.	4

Table 28. Subtopics Identified from the Focus Group Data

Source: Own illustration

Interaction. This theme relates to SPAT being able to build up a dialogue with the students. Several students mentioned that using the SPAT while completing their assignments felt like having an interaction with a peer or a tutor. Students appreciated that the SPAT listened to them and adapted its answers accordingly. Some students mentioned that they like to receive challenging questions and hints from the SPAT; it helped them to think of the next solution steps. Other students perceived it as more exciting compared to “business as usual” paper-based learning materials. For example, one student commented: “I liked her [Alexa]. It felt like I was talking to a teacher. She responded immediately and also asked me challenging questions that helped me with the next steps.” Moreover, several students mentioned that they liked the way the SPAT was helping them.

Specifically, they said that they appreciated thinking on their own about the solution first and that they could control when they received help. For example, one student commented: “It was nice that Alexa was waiting until I asked her for help. That helped me to first think on my own and only receive hints when I want.” Moreover, some students also mentioned that they liked saying their solutions out loud. While hearing themselves talk, they came up with new solution ideas more easily. For example, one student commented: “I liked speaking the answers out loud and not writing them down. When I hear myself, I get new ideas.”

Usage Behavior. This theme relates to the different ways students used the SPAT compared to paper-based learning materials. Most of the students used the SPAT on their smartphone rather than their standalone device and mentioned that it was really easy and fun to access Alexa similar to Google’s Assistant or Apple’s Siri. For example, one student commented: “It was like speaking with Siri. Just like having your personal tutor always in your pocket anytime and anywhere you want.” The easy access to the learning tutor on smartphones and also other devices (e.g., Amazon’s standalone device Amazon Echo Dot) led to a great variety of different learning locations. A handful of students indicated that they did their tasks in other places than usual. For example, one student commented: “When I was lying on my couch, I talked to Alexa a few times, too.” However, a few students also mentioned some areas where SPAT were not very functional. For example, when students were in public, they seldom used Alexa because they felt uncomfortable talking to an SPA in public. One student commented: “First, I wanted to conduct the tasks in the train. But then I decided to do something else; speaking with my smartphone in front of others felt weird to me”.

Individualization. The third topic relates to the capacity of SPAT to recognize students’ individual characteristics. Some students mentioned that they appreciated that Alexa was able to adapt her answers to their utterances and prompted them individually until they were able to find their own solution. For example, one student commented: “It was helpful that Alexa recognized how she can help me discover the solution.” Moreover, some students mentioned that they liked that the SPAT remembered the status of a partly-finished task so they could continue working from there next time. However, some other students wished that the SPAT could recognize where students have their biggest weaknesses. For example, one student commented: “It would be great if she [Alexa] can remember our mistakes and then concentrate on helping us with that.”

5.2.7 Discussion

The aims of the current study were to investigate the effect of conversational instruction (DP2) incorporated in SPAT on high school and vocational school students’ problem-solving skills and how these interactions impact their learning processes. We observed that the students in the experimental groups had higher increases in their levels of problem-solving skills and we discovered changes in how they learned. With the help of conversationally instructive SPAT, students

were able to build up individual interactions and receive individual support on devices they use every day. These quantitative and qualitative data indicate a positive effect of the DP “conversational instruction”. Interestingly, this effect was even stronger in vocational business schools where students’ characteristics are different. This study expands prior research around SPAT that used mainly explorative approaches (Dousay and Hall 2018).

Although the quantitative data of the rather small sample do not allow more detailed quantitative analyses, the qualitative data indicates reasons for the positive effect of conversationally instructive SPAT on problem-solving skills. First of all, SPAT is able to build up multi-turn dialogs with students that help to trigger students’ interactive learning behavior. According to the constructivist learning paradigm and the ICAP-framework (Chi and Wylie 2014), interactive learning behavior occurs when students discuss the learning content with others where both make a similar contribution (see section 2.2.2). This behavior can be considered as the gold standard leading to better learning outcomes. Our findings suggest that conversationally instructive SPAT might replace these discussion partners to some extent. This can also be shown by our qualitative findings indicating that students perceive the interaction with SPAT as similar to interactions with human tutors (theme 1 in the focus group discussion). This reawakens the discussion about the role of SPAT and other AI systems in education. Should AI systems be an *educator-replacing* or an *educator-assisting* technology? The role of AI used only for automation in education systems is questioned (Baker 2016). Cukurova (2019) argues that AI systems should be considered as on a continuum with regard to the extent they are decoupled from humans, rather than only an approach to provide full automation. Indeed, we argue that SPAT can be positioned on both sides of the continuum. In our setting, SPAT have served mainly as educator-replacing to aid educators in settings where they cannot provide individual support due to financial and organizational reasons. SPAT could therefore be used mainly in isolation from educators. In addition, SPAT could also be used to support the decision-making processes of educators. When students interact with SPAT, a great amount of student data can be collected, such as the current level of knowledge, frequency of exercises, motivation at individual, and collective levels. This student data can help educators to introduce new measures (e.g., to explain ununderstood contents again, to offer low-achievers individual help). In my view, the question about the role of AI systems in education is less about an either-or but more about both: i.e., replacing educators where they cannot offer help and informing educators’ decision processes. Moreover, the qualitative results indicate that students liked how the SPAT fostered their understanding until they found the answer on their own. This indicates the positive effect of dynamic scaffolds and is a big difference to “business-as-usual” paper-based learning materials, where there is no interaction possible and scaffolds are static. It seems that students enjoyed having their personal tutor on their side and knowing that they can receive individual support when they cannot get any further on their own. The SPAT then supported them in constructing knowledge on their own. This

helped students to better internalize the learning materials and gain their problem-solving skills.

Second, what could also contribute to the positive relationship between using conversationally instructive SPAT and students' problem-solving skills might be that students can change their learning behavior (theme 2). In particular, students can change their learning places and learning times. Similar results can be confirmed by Taylor (2006), who stated that using smartphones as learning assistants changed learning contexts, for instance in terms of ergonomics (user posture, lighting, background noise), social context, and demands of users' attention. The different learning places and times might lead to higher amounts of learning time and, finally, better skill-development. On the other hand, students also mentioned that they felt weird using SPAT in public, which is also confirmed by findings from Moorthy and Vu (2015). Their results showed that participants preferred using SPAT in a private location (e.g., their home) due to social acceptability.

Third, one other reason for the superiority of the SPAT compared to paper-based learning materials might be that SPAT is able to provide more dynamic support. In specific, students highlighted that the SPAT allowed them to learn at their own pace, receiving help whenever they wanted it. Learning at their own pace motivates students and gives them the feeling of working on their own academic progress (Chen 2008). These effects of receiving individual help whenever they want can also be confirmed by several other research papers in the area of personalized learning (Ammar et al. 2010; Song et al. 2012). Students did however mention that they wished that SPAT could detect individual students' characteristics and in particular track their weaknesses. This confirms the research effort in the area of computer tutoring systems trying to capture students' characteristics and adapt the systems accordingly. For example, Ammar et al. (2010) designed a computer tutoring system equipped with emotional management capabilities, which can capture student's emotions during learning and respond accordingly.

This work contributes to two different research areas in educational technology. First of all, I contribute to computer tutoring research by providing much-needed empirical evidence on the effect of conversational instruction with a new kind of ITS (SPAT) on students' long-term skill development compared to paper-based learning materials, especially in the field of business and law. To the best of our knowledge, empirical evidence for this technology in this area is still missing. Second, I contribute to technology-mediated learning by arguing that interactions with a new kind of IT have the potential to change students' learning processes, resulting in increased levels of skill development. The qualitative results assume that this new technology might be able to offer instructions in a more natural and sophisticated way. Since student-teacher ratios are increasing in today's learning scenarios such as university mass lectures and online courses, SPAs could be very useful. In regard to its practical implications, this study exemplarily showed educators how to build and integrate SPAs in an existing "business as usual" learning environment.

5.2.8 Limitations and Future Research

The present study is not without limitations. First, we used a field quasi-experiment design to examine the impact of SPAT on students' problem-solving skills. With this treatment design, pre-treatment group differences between the experimental group and the control group may confound post-treatment outcomes. We tried to address this point by collecting pre-experiment data and using ANOVAs to check if experimental and control groups are similar. Moreover, both classes in each school had the same teacher with the same learning goals and using the same teaching methods. Nevertheless, future experimental research may be able to confirm and extend our results.

Second, the sample size of the two field quasi-experiments ($n=45$, 2 classes per school) can be considered as rather small for using an ANCOVA. However, the statistics showed a medium effect (in the high school) and a large effect (in the vocational business school) between these two groups, which indicates that there is a relationship between SPAT usage and the increase of problem-solving skills. However, this small sample size does not allow further investigations regarding subsets of the sample. For example, it would be particularly interesting to analyze low achiever's engagement with SPAT compared to high achievers or those with low and high metacognitive skills to self-regulate their learning process. In addition, it may be revealing to observe students using these SPAT and derive interaction patterns from these data. These research questions cannot be answered with this study but provide exciting avenues for future research. Furthermore, I applied a mixed-method approach to partially compensate for the rather small sample size (Venkatesh et al. 2013). It would be interesting to see if studies with larger sample sizes are able to confirm our effect. Third, the focus group consisted only of participants of the second experiment in the vocational school. I cannot completely exclude the possibility that the statements made by participants of the first experiment (high school) would differ. Fourth, I conducted the SPAT experiment over a relatively brief period under positive conditions. Such studies tend to produce better outcomes due to novelty effects and hyperattention to experiment details (Cheung and Slavin 2013). For the current study, novelty effects may be small, given the large percentage of participants reporting already high usage of SPAT on their smartphones (approximately 40% used SPAT every day on their smartphones and another 40% used it weekly; e.g., Siri). Nevertheless, using SPAT in the context of homework was something completely new for many students, which raises the question of the extent to which these systems can be used in the long term. It would be interesting to see in long-term studies whether these students also show increased problem-solving skills after 3-4 months or even longer. Another exciting question would be to investigate whether the time of use of these devices increases or decreases over the course of a whole semester. This study cannot provide answers to these questions, but they offer interesting avenues for future research. Finally, this study also raises some ethical questions about data security and the potential benefit of one group and not the other. Students put their information on Amazon's ecosystem, which might

be a risk for some students. While it is hard to control student behavior, I ethically approved the study and educated students about the experiment and the corresponding risks at the beginning of the study. For future research, it would be interesting to see how SPAT can be implemented in-class to improve students' skill development. Moreover, future research should focus on individual differences in students' characteristics that might influence learning processes with SPAT.

5.2.9 Conclusion and Next Steps for the Course of the Dissertation

This study investigated how the DP2 "conversational instruction" incorporated into SPAT influence students learning process and metacognitive skill development in a blended learning environment.

In particular, I investigated whether SPAT could increase students' problem-solving skills. Within two experiments in two different school types (high school and vocational school), I was able to show that the use of SPAT over a period of five weeks have a positive effect on skill development, more precisely the development of problem-solving skills. Second, I examined how the use of SPAT changes students' learning processes. Students showed a more interactive learning behavior and used the SPAT in a different way compared to traditional learning aids. Moreover, students appreciated that they received individualized support from the SPAT. The study provides empirical evidence for SPAT and offers first insights into how this technology can change students learning processes. The findings of this study contribute to computer tutoring and TML research. This study is one of two studies to answer RQ2a of the dissertation. With the help of these two studies, I indicated the usefulness of human-like (DP1) and conversationally instructive (DP2) SPAT in blended learning environments. On top of that, there are three DPs (dynamic scaffolding, engaging dialog flow, accessibility) that remain to be investigated. Moreover, I did not consider pure online learning environments so far. Thus, the section 5.3 continues with investigating DP3 (dynamic scaffolding) and DP4 (engaging dialog flow) in online learning environments to answer RQ2b.

5.3 Experimental Evaluation of Smart Personal Assistant Technology included in an Online Video Lecture to Increase Student Engagement and Outcome

5.3.1 Introduction

RQ2b investigates DP3 (dynamic scaffolding) and DP4 (engaging dialog flow) in pure online environments to cover both types of learning environments:

RQ2b: How can dynamic scaffolding and an engaging dialog flow incorporated in Smart Personal Assistant Technology improve students' learning success in online learning environments?

The work that seeks to answer to this question is again divided into two studies. Study 1 is a laboratory experiment, in which I look at DP3 in a controlled environment before investigating DP4 in a real online learning environment.

Enrollment in online courses has grown rapidly (Seaman et al. 2018). In 2017, 33.1% of students worldwide took at least one course online, compared to 24.8% in 2012 (Lederman 2018). With this development, the amount of online educational content is rapidly increasing, particularly in the form of online video lectures (Kizilcec et al. 2014). However, designers of these learning materials are faced with a challenge: educators are ill-equipped at eliciting individualized, meaningful interactions with their students (Song et al. 2017). This is problematic, since we know from learning theory that we learn best when we interact socially with others through meaningful interactions (Kim 2001). Meaningful interactions, such as scaffolding dialogs between an educator and a student, can have a significant effect on learning outcomes, including students' information retention (Ibrahim and Al-Shara 2007) and their ability to apply the new knowledge to solve novel problems—commonly referred to as students' transfer ability (Cortright et al. 2005).

Integrating scaffolding in computer tutoring systems has proven its value, as document in past research (Albacete et al. 2018; Duffy and Azevedo 2015). However, the integration of dynamic scaffolding into SPAT can boost this development further and help these systems to interact with students in a more naturalistic and sophisticated way. To investigate these aspects, I present a study on Sara, a web-based SPAT that can be layered on top of already existing online video lectures. Sara provides voice- and text-based scaffolds to students and thereby helps them to remember concepts better (information retention) and to use the gained knowledge to solve novel problems (transfer ability). At predefined times, Sara interrupts an online video lecture and actively intervenes in the learning process. Sara asks questions and offers detailed explanations, providing scaffolding questions if the student's initial answer is incorrect. This dynamic scaffolding strategy is inspired by previous studies in the field of scaffolding and ITS (Albacete et al. 2018; Ferschke et al. 2015). To determine the true impact of Sara's

dynamic scaffolding approach on students' information retention and transfer ability, I tested it in a 2 x 2 lab experiment with 182 students. I compared Sara with three other types of SPAT that are often used today: (a) a voice-based, statically scaffolding SPAT; (b) a text-based, dynamically scaffolding SPAT; and (c) a text-based, statically scaffolding SPAT. The findings suggest that dynamically scaffolding SPAT is able to increase students' information retention but also helps them to apply the acquired knowledge to novel problems.

I build on previous findings that have investigated the beneficial effects of SPAT in education (Ruan et al. 2019) and online education in particular (Song et al. 2017). Following this line of research, the study highlights the importance of including scaffolding and voice-based SPAT in the context of online video lectures. In contrast to the work of Litman et al. (2006), my paper focuses on employing voice-based dynamically scaffolding SPAT during video instruction when students learn about new concepts for the first time. This is different to scaffolding during the exercise phase when students solve a task. With the help of dynamically scaffolding and voice-based SPAT, static online video lectures may move closer to the gold standard of educator–student interactions. This effect may be even stronger when the SPAT is voice-based. To the best of my knowledge, this is the first study that has designed a voice- and text-based SPAT with a dynamic scaffolding mechanism in an online video lecture context, where the standard method of instruction is solely online videos that prevent students from actively thinking and deepening their understanding of concepts. Moreover, I empirically evaluated and rigorously compared dynamic scaffolding with other types of SPAT. This enlightens software designers and educators on how SPAT can be used in their online education courses.

5.3.2 Related Work and Hypotheses Development

Among other things, the theory of scaffolding states that when students strive to acquire new knowledge, they need individualized support that falls within their individual ZPD. ZPD represents the potential distance the student could reach with the help of a more knowledgeable other (Vygotsky 1978). As the students advance and become more independent in their thinking, this support can gradually be taken away. Dynamic scaffolding is a term used to explain the interaction between students and their educators and is a process that enables a novice to achieve a goal or an objective that would otherwise be unattainable without assistance (Wood et al. 1976). The main goal of the educator is to offer scaffolds, such as questions and hints, within students' individual ZPD. Dynamic scaffolding is not one-way but is an interactive and reciprocal process between the student and the educator (Bull et al. 1999). In the context of my study, this means that dynamic scaffolds have to be adaptive and personalized rather than one-size-fits-all video lectures. Dynamic scaffolding during instruction improves students' ability to remember concepts (Wood et al. 1976). However, in online learning, it is difficult for an educator to offer dynamic scaffolds to every student (Rietsche et al. 2018). Dynamically scaffolding SPAT might be able to imitate the educator's

scaffolding dialogs to some extent resulting in a meaningful learning process. I therefore propose the following hypothesis:

H1a: Students interacting with dynamically scaffolding SPAT show higher levels of information retention compared to those interacting with statically scaffolding SPAT.

Past research indicates that dynamic scaffolding not only helps students to remember learned concepts but also to increase their transfer ability (Applebee and Langer 1983). This has been shown both in problem-solving tasks in a computer course (Wang et al. 2015) and in solving novel problems in physics (Murphy and Messer 2000). Wang et al. (Wang et al. 2015) created iTutor to help students to learn basic computer skills. Their results indicated that students in the iTutor group experienced better learning effectiveness than those in the control group. Moreover, Murphy and Messer (2000) were able to prove that students can better transfer their knowledge to novel problems in a physics task when receiving scaffolds compared to when they are working in a group or working alone. I therefore propose the following hypothesis:

H1b: Students interacting with dynamically scaffolding SPAT show higher levels of transfer ability compared to those interacting with statically scaffolding SPAT.

Learner engagement can be described as the holy grail of learning (Halverson and Graham 2019). Past literature illustrates that learning success occurs when learners are engaged in their own learning (Sinatra et al. 2015). Learner engagement is a measure that reflects the quantity and quality of a learner's involvement with their learning material (Halverson and Graham 2019). Learners are engaged when they are active in their learning, eager to participate, willing to expend effort, motivated, and inspired (Sinatra et al. 2015). Active learning requires learners to engage cognitively and meaningfully with the tasks they are dealing with. When learning actively, learners think about their learning material in depth rather than just passively receiving it. When learners are in a passive engagement mode, they consume information passively without performing any physical activity—for example, watching a tutorial without doing anything else other than watching. In contrast, an active engagement mode means that a physical activity is carried out. An example of this is the marking of text passages while reading. A constructive engagement mode goes one step further. The learner creates new output that goes beyond the subject matter, such as writing notes in their own words during the lecture. An interactive engagement mode can be considered as the gold standard of learning, a situation in which both partners make similar contributions to the dialog. An example of this is a scaffolding dialog between an educator and a learner (Chi and Wylie 2014).

Already implemented SPAT in online learning may enable learners to adopt a constructive engagement mode by helping them to reflect on the learning material (e.g., summarizing the content of a video). In contrast to this, I argue that dynam-

ically scaffolding SPAT might be able to trigger learners' interactive learning engagement mode by building up a scaffolding dialog with learners and not ignoring a partner's contribution (Chi and Wylie 2014). This kind of SPAT is able to adapt to the contributions of the learner where the dialog is primarily constructive and a sufficient degree of turn-taking occurs. I therefore propose the following hypothesis:

H2. Dynamically scaffolding SPAT built into online video lectures positively influences levels of learner engagement compared to statically scaffolding SPAT.

Voice-based instruction has benefits over a text-based approach. SPAT's spoken output can elicit a warmer attitude among users toward the agent and lead to a richer use of language (Novielli et al. 2010). Students also achieve a higher content coverage in the spoken condition and a better adjustment of communication when speaking compared to communicating via text with a SPAT (D'Mello et al. 2010). Moreover, de Kock et al. (2016) found that students who worked with audio/visual hints improved their information retention and solved more problems correctly in the transfer test compared to those who received text-based hints. The research stream of the think-aloud method, where students speak their knowledge out loud, confirms the positive effect of voice over text (Bannert and Mengelkamp 2008). I therefore propose the following hypotheses:

H3a. The scaffolding effect will be stronger for information retention when the SPAT is voice- and text-based rather than textual.

H3b. The scaffolding effect will be stronger for transfer ability when the SPAT is voice- and text-based rather than textual.

5.3.3 Sara, a dynamically scaffolding SPAT

To investigate the hypotheses presented in Subsection 5.3.2, I designed a SPAT that I named Sara. To implement the DP3 (dynamic scaffolding), Sara includes a *main- and sub-dialog* interaction logic, which is derived from the prominent scaffolding strategy called *directed lines of reasoning*, in which educators ask a series of directive questions to help the student dive deeper into the topic (Mills et al. 2004). If the students' answers are correct, the educator advances to the next question. Otherwise, the educator tries to scaffold the students' understanding with additional questions that meet the students' individual ZPD for the targeted concepts (Albacete et al. 2018). Figure 23 illustrates this DP with the help of an example interaction between a student and Sara. The video depicted in the interaction deals with learning how to create a conditional execution in Python. I implemented the dynamically scaffolding logic in the prototype as follows: Sara interrupts the video lecture after a specific time (i.e., a subchapter) and asks each student a series of comprehension questions to repeat the just seen and heard content. Sara is able to detect three different kinds of answer: *correct*, *wrong*, and *don't know*. When Sara detects a correct answer, she continues with the main dialog (left side of Figure 23).

When Sara detects a wrong answer or when the student does not know how to reply, Sara opens a sub-dialog (right side) where she provides feedback, an explanation, and scaffolding questions within the student's ZPD that help them to close their knowledge gaps and find the answer to the initial question of the main dialog. After asking the scaffolding question, Sara explains the answer and finally asks the main question again. The students have another chance to answer the question before Sara explains the answer and continues with the next question of the main dialog. Using this mechanism, Sara tries to imitate educators' scaffolding behavior when interacting with their students. Finally, after Sara has interacted with the students for several minutes, the video lecture continues with the next subchapter and the SPAT falls silent. After the next section or segment, the SPAT interrupts the video lecture again and commences with the next dialog.

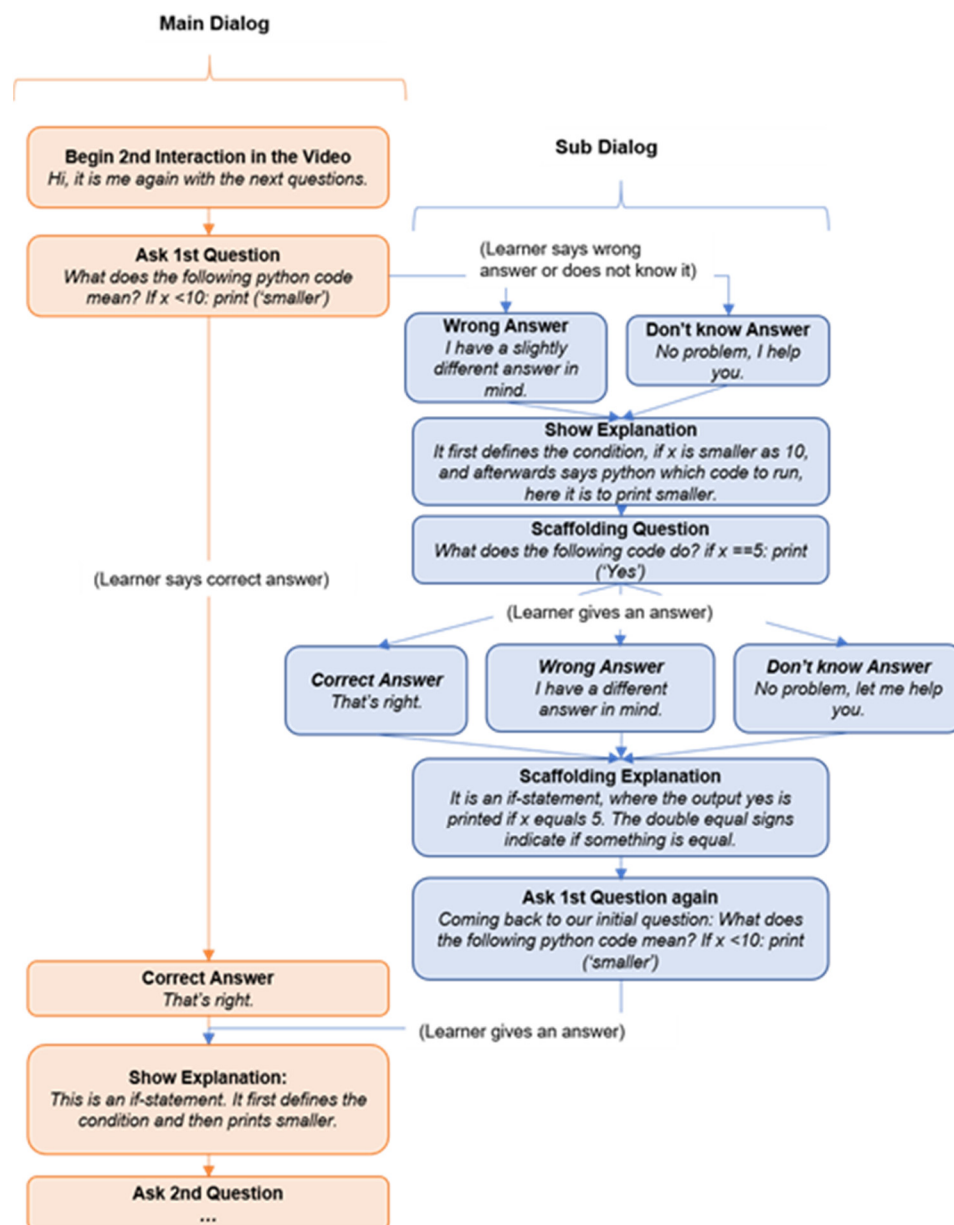


Figure 23. Main- and Sub-dialog Mechanism of SPAT

Source: Own illustration

To provide dynamic scaffolds, the system has to include mechanisms that conduct some form of speech analysis that helps in deciding whether a student's answer is right or wrong or if they do not know how to respond to the question (Graesser et al. 2017b). To implement this principle, there are different ways in which SPAT can detect the kind of answer given. In many cases, simple pattern matching is conducted to compare the inserted answer to one or multiple optimal solutions (Latham et al. 2010). However, in open-topic natural language communication, as used in this study, defining all optimal solutions is difficult. There is no single expression that describes the correct solution—rather, different descriptions can show a correct solution. Hence, I used a more sophisticated solution based on a pre-trained model. I used the NLP.js framework (AXA Group Operations 2018) for the intent classification, as this has been shown to be one of the best NLP libraries in recent benchmarks (Git Hub 2019). NLP.js is publicly available under <https://github.com/axa-group/nlp.js>. Unlike other cloud-based services, it enables all data to be stored and processed in our own infrastructure, which is required in many educational settings (due to privacy regulations). For each question Sara asked the students (see Figure 23), I trained one NLP model using a training set consisting of multiple possible *correct*, *wrong*, and *don't know* answers provided by seven students' and three researchers' pretests resulting in a set of approximately 800 single statements. Once the student answers a question, the voice-based answer is recorded, automatically transcribed to text using the HTML 5's web Speech API (W3C Community and Business Groups 2019), and the resulting textual data is then sent to the NLP server, where I apply the NLP models to classify the student's answers using the pre-trained models. Using the results of this intent classification step (i.e., answer was correct, wrong, or student does not know the answer), I decide whether Sara should continue with the main dialog (correct answer, see Figure 23 left side) or open the sub-dialog (wrong or don't know answer, see Figure 23 right side).

To implement voice recognition, Sara starts talking to the student via voice based on the HTML 5's Speech Synthesis API (W3C Community and Business Groups 2019). Sara generates a greeting statement via voice and asks the student if they have understood the initial instructions. Additionally, Sara's spoken text is displayed in the chat window directly after she finishes her utterances (see Figure 24).

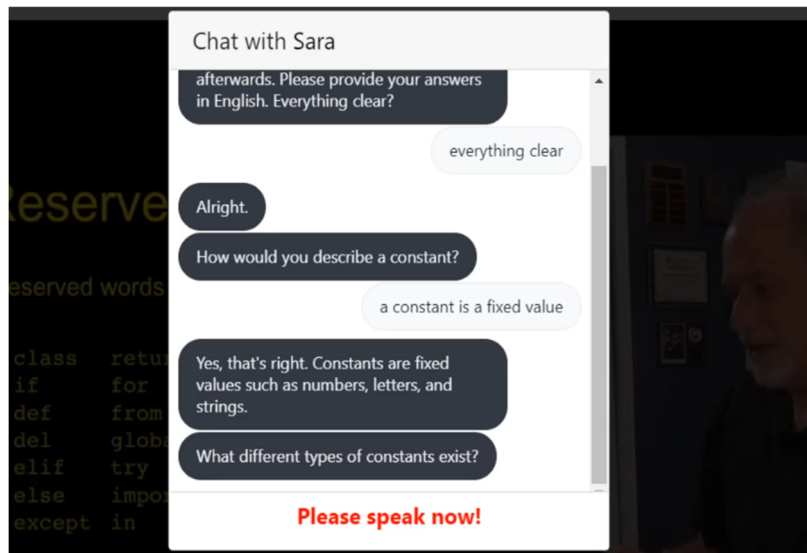


Figure 24. Screenshot of SPAT During an Online Video Lecture, Shown on Top of the Video Screen.

Source: Own illustration

As shown in Figure 24, I integrated Sara in a web-based video player. While the video lecture is playing, Sara is hidden. When a certain topic has been completed in the video lecture, Sara appears in front of it, and the video is paused. The student can then interact with Sara based on her scaffolding strategy to deepen their understanding of the learning content just seen. After the interaction is completed, the video continues with the next subchapter until Sara appears again. I showed our students two different video lectures. The timeline of video 1 includes the interaction and is displayed in Figure 25. I argue that with this educator-like type of interaction, a better learning success can be achieved.

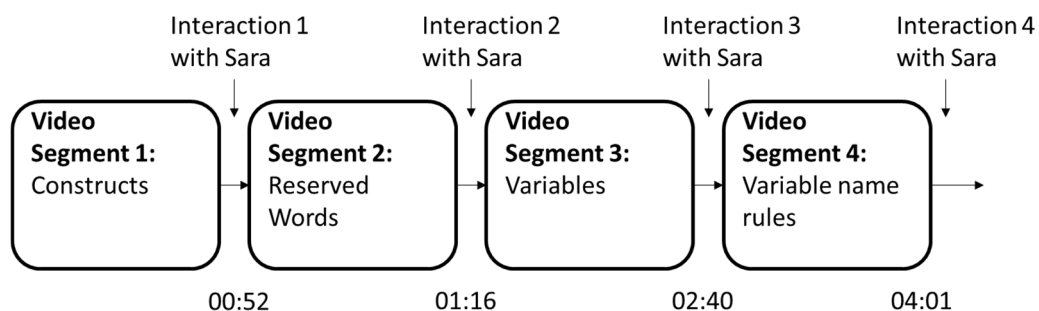


Figure 25. Timeline of Video 1 and its Interaction Points.

Source: Own illustration

5.3.4 Experimental Evaluation of Sara

The designed experiment tests our hypotheses and tests Sara through a rigorous multi-comparison of different SPAT types. I used a 2 (dynamic scaffolding, static scaffolding) x 2 (textual, voice- and text-based) x 2 (repeated measure, 2 random treatments per participant) design. I decided to use a repeated-measure design

that involved showing students two videos, to make sure that the results were not restricted to the content of only one video. I compared Sara to three other types of SPAT that are often implemented in learning settings. The SPAT comparisons were between (1) dynamically scaffolding SPAT (Sara); (2) a voice- and text-based statically scaffolding SPAT; (3) a textual scaffolding SPAT; and (4) a textual, non-scaffolding SPAT. Moreover, I included a control group in which students watched the videos only. Students watched two online video lectures and interacted with one SPAT type per video. The SPAT interactions took place four times per video (except for the CG, where students simply watched the video lecture). The students received the treatments and the videos in a random order.

Figure 26 compares a scaffolding SPAT and a non-scaffolding SPAT. The non-scaffolding SPAT offered standardized, one-size-fits-all hints, where students were able to click or say “next”. These kinds of SPAT have often been implemented in online learning settings in recent years (Kulik and Fletcher 2016). The hints included the same information as the main dialog of the scaffolding SPAT (see red rectangles in Figure 4). The only difference was that the scaffolding SPAT used questions to talk with the students and was able to open a scaffolding sub-dialog when students provided a *wrong* or *don’t know* answer. The voice- and text-based SPAT allowed students to provide their answers via voice. Moreover, Sara provided her questions and explanations via voice and displayed the text on the screen afterwards (multi-channel, DP3). The text-based SPAT allowed students to type in answers via the keyboard, an input mode that is often seen in SPAT in online education (Ruan et al. 2019). Our DVs were the students’ information retention and transfer ability levels. I used a pretest and posttest design to measure the differences in the posttest scores of the DVs across the SPAT groups and the control group, including the pretest scores, learning channel preference, and cognitive load as covariates (i.e., treatment scores).

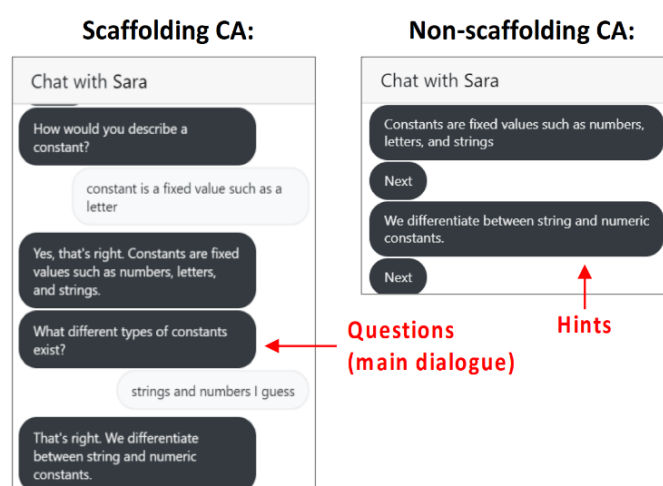


Figure 26. Dynamically Scaffolding vs. Statically Scaffolding SPAT Interaction. Left: Students are Asked for Answers. Right: Students see Hints and can Advance by Clicking or Saying "Next".

Source: Own illustration

5.3.4.1 Participants

I recruited undergraduate and graduate students ($n = 182$, 74 female, 108 male, aged 18 to 35) from the participant pool of a European business university. Table 29 shows the sample's characteristics across the four TGs and the control group. I randomly assigned the students to one of the four treatment groups or the control group. They received USD 20 as a baseline and an additional USD 10 depending on their gain scores (difference between the pretest and posttest). The randomization was successful, since independent samples t -tests revealed that all the SPAT groups were similar in terms of *age*, *gender*, *experience with SPAT*, *personal innovativeness*, *language level*, and *nationality* ($p > .05$).

	Non-scaffolding, textual	Scaffolding, textual	Non-scaffolding, voice-based	Sara	Control group (video only)
Sample size	37	36	37	37	35
Gender (M/F)	23/14	21/15	22/15	21/15	23/12
Avg. age	22.4	22.8	23.1	23.0	22.2
Prev. exp. with SPAT (usage per week)	3.6	3.7	3.9	3.5	3.9
Personal innovativeness (out of 7)	4.7	4.8	4.9	4.8	4.8
Language level (out of 7, 7 = proficient)	5.4	5.6	5.2	5.3	5.5

Table 29. Sample's Characteristics

Source: Own illustration

5.3.5 Materials

5.3.5.1 Videos

Each student watched two beginner-level videos about Python programming in a randomly assigned order. I chose these videos for two reasons. First, I wanted to make sure that the students had little previous knowledge of the topic. Since the study participants were business school students who do not attend a programming class, the topic was suitable. Second, I wanted to ensure that the videos were representative in terms of quality, typicality, and video format. Both videos were from a popular Python course taught by Charles R. Severance at the University of Michigan School of Information, publicly available under

<https://www.py4e.com/>. Being about Python programming, the topic was representative of online learning material in which the amount of technically oriented learning content has increased significantly in recent years (Lepp et al. 2017). The video format was web-based and picture-in-picture, which is very often used in MOOCs (Kizilcec et al. 2014). The videos I chose for the experiment were about constants and variables (video 1; length 4 min 52 s without SPAT interaction) and conditional execution (video 2; length 4 min 48 s without SPAT interaction). The videos presenting this content were largely independent of each other. Each video started with a 10 second resting period (black screen with white text stating “Resting Period”). I asked the students in the post-survey to rate the difficulty of the two videos on a scale from 1 to 7 to ensure that the videos were equally difficult for all the students and that possible effects could not be attributed to differences in the difficulty of the videos. The students’ ratings of the videos’ perceived difficulty were on the same level (Video 1: 5.43 out of 7, Video 2: 5.63 out of 7, $p > .05$).

5.3.6 Tests and Questionnaires

The pre-survey contained questions about students’ preferences regarding visual and auditive learning styles (obtained from (O’Brien 1989)). I included this variable as a covariate in my data analysis to factor out the possibility that differences in the SPAT groups might be attributable to different learning styles. The pre-stimulus test contained six multiple choice questions (three questions per video) and two open problem questions (one question per video) to measure students’ relevant pre-knowledge level on programming with Python. The post-stimulus test followed directly after the stimulus and contained exactly the same questions as the pre-stimulus test. All the test questions were presented in a random order. The six multiple choice questions measured *information retention* and the two open questions measured *transfer ability*. The multiple-choice questions related to information retention, asking the students to reproduce content from the video. The open questions required students to independently apply the acquired knowledge to a novel problem. For example, in the multiple-choice question section, I asked students to mark the statement that best corresponds to the description of a constant. In the transfer question part, the students actively solved a small programming exercise. For example, I asked them to define and name two new variables, assigned with the tags “Martin” and “7”, and to print them as output. The post-stimulus survey contained items concerning cognitive load (from Lin et al. (2013)), perceived difficulty of the videos, personal innovativeness (from Raaij and Schepers (2008)), language proficiency, personal experience with SPAT, age, gender, and nationality, plus one open question about how the students experienced the SPAT. I used cognitive load scores as covariates in our data analysis to control for differences in cognitive load when using different SPAT types.

5.3.7 Procedure

The experimental procedure is depicted in Figure 27. The students carried out the experiment individually and one after the other. After obtaining consent, the researcher presented the experiment to the student and the student started with the pre-survey and pre-stimulus test in room 1, without the presence of the researcher. After conducting the pre-survey and pre-stimulus test, the researcher returned and escorted the student into room 2, which contained a chair facing a computer monitor. The researcher instructed the student to pay attention to the two lectures, as there would be a short quiz on the material afterwards. The researcher then departed, after which the participant watched the stimulus without pauses and was not allowed to take notes. Once the two video lectures had ended, the researcher escorted the participant to room 1 again to conduct the post-stimulus test and post-survey. After completion, the participant was asked to guess the purpose of the study and was then debriefed on the experiment. Most of the participants thought the study concerned how to effectively learn programming with Python.

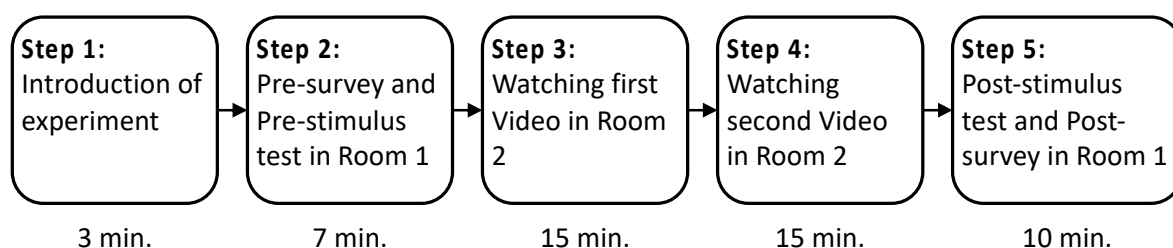


Figure 27. Experimental Procedure

Source: Own illustration

5.3.8 Measurement and Data Analysis

5.3.8.1 Learning Outcome

For measuring the DVs—*information retention* and *transfer ability*—two experienced raters evaluated the results using an evaluation framework that was developed collectively in a workshop within the research team. In the multiple-choice question section, the evaluation framework provided the correct answer and specified how many points to give. For each correctly answered question, I awarded one point. In the open transfer question section, the evaluation framework specified exactly what points were to be awarded for what and how many points should be deducted for common mistakes (e.g., two points given for the correct syntax of the if-statement, two points given for the correct semantics, half a point subtracted if a colon was missing, etc.).

The post-stimulus test results were rated individually, blinded (i.e., raters do not know the SPAT group), and assessed independently of each other. Furthermore, for the transfer open question section, I checked for interrater agreement with the help of a Pearson correlation. Interrater agreement gives a score showing the

amount of homogeneity or consensus in the ratings (Fleiss et al. 1981). The correlation resulted in a satisfactory result (interrater agreement = 0.94, $p > .05$). To identify differences between groups who had interacted with the SPAT types, I first conducted ANCOVAs to ascertain whether differences existed across the groups, including the control group. ANCOVAs help to identify whether there are significant differences in variances between the groups while controlling for covariates. I did not use ANOVA, because I needed to factor out errors that were introduced by covariates and ANOVA masks the true relationship between the type of SPAT and the DVs. I thus included the pretest scores, cognitive load, and the learning channel preference (auditive or visual) as covariates. All these covariates have the potential to disturb the true effect between the SPAT and learning outcomes. I also verified that the data meets the assumptions of an ANCOVA (normality, equality of slopes of the covariates and the outcome variable, equality of the groups and the covariates, and homogeneity of variance) with t -tests and a Levene's test (Levene 1960). All assumptions were met.

After conducting the ANCOVAs, I investigated the patterns and comparisons between specific groups in order to test hypotheses H1a and H2b. Although it is otherwise a powerful tool, ANCOVA cannot provide such results. Hence, I conducted a post-hoc analysis using Tukey's Test (Tukey 1949) to test my hypotheses. Tukey's test compares the means of all treatments to the mean of every other treatment and is considered the best available method in cases where confidence intervals are desired or where sample sizes are unequal. To analyze the interaction effect between *static scaffolding/dynamically scaffolding* and *text-based/text-* and *voice-based*, I used factorial ANCOVAs with both factors as independent variables to test H3. I calculated partial eta squared as a measure of the strength of an effect (0.01 = small, 0.06 = medium, 0.14 = large, (Cohen 1973)).

5.3.8.2 Learner Engagement

The concept and measurement of learner engagement has been widely discussed in educational psychology (Shernoff et al. 2014). The types of measurements can be roughly categorized into macro-level (e.g., learner's self reportings, observations) and micro-level measurements (e.g., neuroIS measurements). Research papers in the area of educational psychology suggest that learner engagement is better measured on a micro-level (Siadaty et al. 2016). Macro-level measurements, such as self-reportings, have the inherent problem of retrospection or biased results. For example, in self-reportings, learners are asked to reflect on the lesson they have just experienced as they consider their responses. Observations, another type of macro-level measurement, eliminates the need for retrospection, because the observations can be made in real time but involve a bias from the observer.

To measure the DV student engagement, I used two common neuroIS measurements: galvanic skin response and facial expression score.

Galvanic Skin Response Score

The galvanic skin response measures a short-term decrease in the electrical resistance of the skin caused by the typical increase in the sympathetic tone during emotional-affective reactions (Francis and Oliver 2018). This leads to an increased secretion of sweat, corresponding to an increase in skin conductance (Lykken and Venables 1971). Since any physiological arousal associated with emotion alters skin conductivity, measurements of electrodermal activity can be used to objectify psychophysiological relationships. Several studies have used galvanic skin response scores to measure student engagement. For example, McNeal et. al's (2014) research measured learner engagement during a geology course using the galvanic skin response score. Before the students were allowed to watch the videos, I attached shimmer™ finger sensors to their index, middle, and ring fingers, which allowed the skin level to be measured throughout their viewing of the video. Once the sensor made skin contact, it automatically began collecting data. The data was then downloaded to the software iMotions™ to measure and analyze the skin conductance levels. Data were collected every 0.125s. The unit of measure for skin conductivity in the International System of Units is Siemens (Boucsein 2013). I plotted the large dataset to gain a first impression of the data and formed an average of the whole galvanic skin response score for both videos. I omitted the scores for the phases when students were interacting with the SPAT, because I aimed to measure the effect of SPAT on student engagement during the phase when the online videos were being watched. I standardized the scores between 0 and 1 to make them comparable with the facial expression scores and computed a final engagement score.

Facial Expression Score

In face coding, human emotions are measured by facial expressions. With facial expression analysis, it is possible to test the impact of content, products, or services that are intended to evoke emotional excitement and facial reactions (Brackett and Katulak 2007). One of the strongest indicators of emotions is facial expression—for example, laughter is associated with certain changes in important facial features. Computer-based facial expression analysis attempts to mimic human coding skills by capturing unfiltered emotional responses to any kind of emotionally appealing content. The facial expressions are captured by a camera and the resulting emotional states are analyzed by automated computer algorithms (iMotions 2019). Once captured, the videos can be imported into software for facial expression analysis. I used the software iMotions™ to gather facial expression measurements. iMotions offers a cumulative facial expression score that is calculated from the weighted sum of the following facial movements: brow raise, brow furrow, nose wrinkle, lip corner depressor, chin raise, lip pucker, lip press, mouth open, lip suck and smile (iMotions 2019). The more often the system detects one of these movements, the higher the facial expression score. I omitted the data for the phases when students were interacting with the SPAT to ensure that data for the phases when they were interacting with SPAT were not

falsifying the scores. I standardized the engagement score between 0 and 1 and computed a final engagement score, where galvanic skin response and face expression are weighted the same. Figure 28 shows an exemplary plot of a student from the dynamically scaffolding TG with time (in minutes) on the x-axis, and galvanic skin response score (GSR) and facial expression score on the y-axis. I have also added the corresponding SPAT interaction phases.

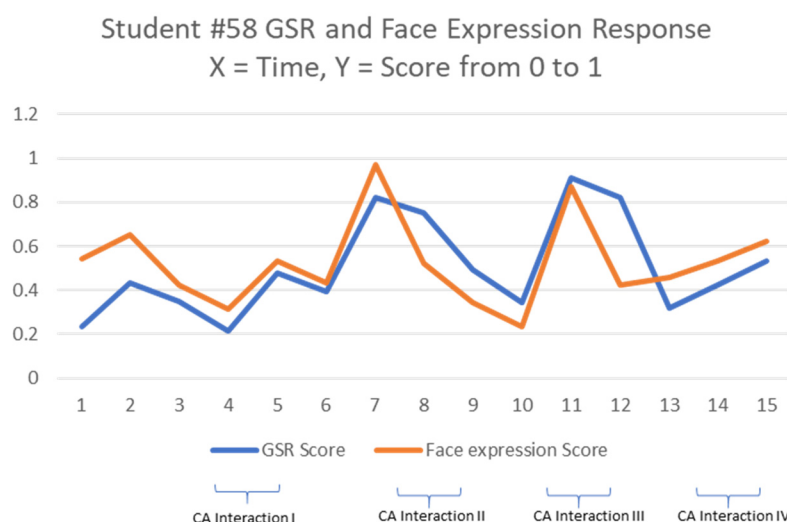


Figure 28. Student #58 GSR and Facial Expression Response for the Dynamically Scaffolding TG.

Source: Own illustration

For data analysis, I first conducted an ANCOVA to identify differences between the groups (including the control group). I then conducted a simple linear regression to test the difference between dynamic and static scaffolding. I conducted a test for normality, homogeneity of variance, and homogeneity of regression slopes to check whether the data met the assumptions for using an ANCOVA. I used the statistic program R as a tool for analysis.

5.3.8.3 Qualitative Data

I used the qualitative data from the open question of the post-survey to gather a better understanding of the relationship between dynamically scaffolding SPAT and student engagement. I asked the students the following question: How did you experience the interaction with the SPAT? I used a thematic analysis to induce topics following the method of Ryan & Bernard (2003). Specifically, I used the keywords-in-context method for this study. With the help of this technique, I analyzed the data in a two-step process. First, I identified key words indicating students' perceptions of the interaction with the SPAT. I systematically searched the corpus of texts to find all instances of the word or phrase. Each time I found a word, I made a copy of it and its immediate context. Second, I identified themes by physically sorting examples with a similar meaning into categories (Ryan and

Bernard 2003). Moreover, I conducted a respondent validation by letting participants review our identified themes (Torrance 2012).

5.3.9 Results

5.3.9.1 Learning Outcomes

The main goal of my study was to investigate the usefulness of DP3 (dynamic scaffolding). Specifically, I tested whether the dynamically scaffolding SPAT Sara is better able to increase students' information retention and transfer ability during online video lectures compared to other often implemented SPAT. Table 30 shows a summary of the pretest and posttest scores of the groups for the two different types of learning outcomes: *information retention* and *transfer ability*. Since I had different maximum points in the tests of the two videos, I normalized all the scores between 0 and 100, where 100 corresponds to the maximum number of points scored at the pretest and post-stimulus test. Descriptive statistics for both *information retention* and *transfer ability* provided similar results: students interacting with Sara showed both the highest level of *information retention* (score = 77.8) and the highest level of *transfer ability* (score = 65.4). These scores were followed by a textual dynamically scaffolding SPAT (73.1 and 54.6), a voice- and text-based statically scaffolding SPAT (65.3 and 51.5), and a textual scaffolding SPAT (61.6 and 50.6). The control group that did not interact with a SPAT had the lowest score in both outcome variables (60.9 and 44.0). The corresponding visualization is depicted in Figure 29.

	Information retention			Transfer ability		
	Pre-test	Post-test	SD post	Pre-test	Post-test	SD post
Non-scaffolding, textual	40.3	61.6	24.6	31.5	50.6	24.3
Scaffolding, textual	39.8	73.1	22.0	26.4	54.6	27.1
Non-scaffolding, voice-based	38.0	65.3	23.2	32.6	51.5	25.7
Sara	34.8	77.8	21.7	26.0	65.4	27.1
Control group (video only)	36.2	60.9	27.2	28.5	44.0	21.9

Table 30. Summary of the Pretest and Posttest Scores

Source: Own illustration

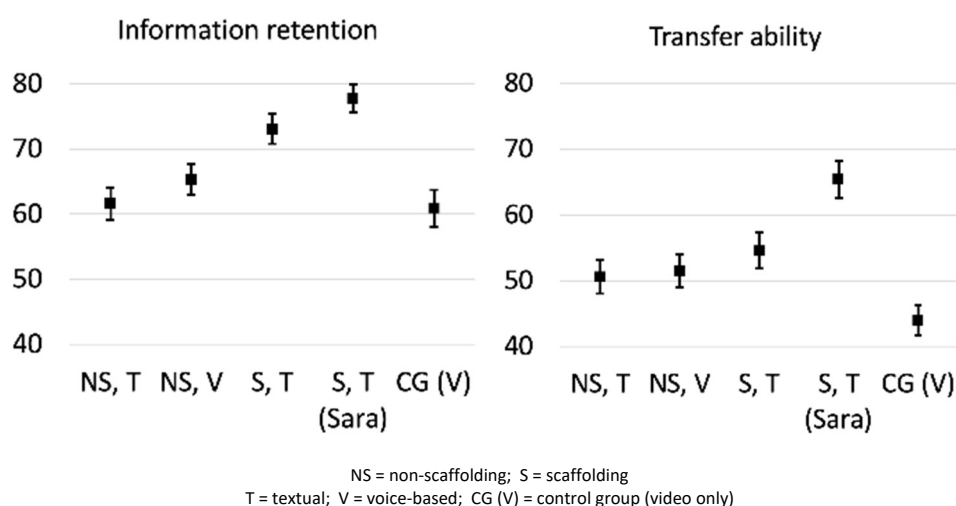


Figure 29. Normalized Post-test Scores for Information Retention and Transfer Ability for Different Conditions. 50% of the Post-test Scores are within the Error Bars.

Source: Own illustration

The ANCOVA test for *information retention* yielded a significant main effect for the dynamically scaffolding SPAT when the pre-test, cognitive load, and the learning channel preference were controlled ($F(4,359) = 29.9, p < .0001^{***}, n = 182$). Partial eta-squared was 0.08, which is considered a medium effect strength (Cohen 1973). Tukey's test revealed that the dynamically scaffolding SPAT was significantly better in supporting information retention compared to statically scaffolding SPAT (*adjusted* $p < .0001^{***}$).

Thus, I **observe support for H1a**. Furthermore, I calculated the intervention selection accuracy (ISA) for the two groups that interacted with the dynamically scaffolding SPAT. I define the ISA in the context of this study by the number of correctly assigned scaffolding sub-dialogs (i.e., each time the classifier tagged a student answer correctly as “wrong” or “don’t know” and thus opened a dynamically scaffolding sub-dialog—see right side of Figure 23) divided by the total number of all sub-dialogs given (Rosé and Vanlehn 2005). The research team tagged the interaction logs and found that in total 51 out of 73 students entered the sub-dialog part of the scaffolding mechanism at least once with 199 sub-dialogs in total (out of 1,168 potentially possible sub-dialogs). In the text-based SPAT group, there were 96 correct interventions out of 102. In the SPAT Sara group, there were 94 out of 97. This leads to an ISA of 95.48%. Although some students of the SPAT Sara group did not enter the dynamically scaffolding sub-dialogs, I argue that this detail does not affect the results and implications, as even the students that did not enter the sub-dialogs interacted with the SPAT during the main dialog questions. I therefore argue that a manipulation has still taken place as a consequence of asking those main questions during the experiment. This interaction might already have stimulated the students to reflect and think about the learning content even though they did not enter the scaffolding sub-dialogs.

Excluding those students would disregard the effect of the main dialog interaction. Consequently, I did not exclude those students from the analysis.

The mean differences and the 95% confidence intervals between the specific groups (control group, statically scaffolding, dynamically scaffolding) are displayed in Figure 30 (left). If an interval does not contain zero, the corresponding means are significantly different. The ANCOVA for *transfer ability* also yielded a significant main effect for the dynamically scaffolding SPAT, again controlling for the pretest, cognitive load, and the learning channel preference ($F(4,359) = 22.3$, $p < .0001^{***}$, $n = 182$). Partial eta-squared was 0.06, yielding again a medium effect strength. The dynamically scaffolding SPAT was significantly better than the statically scaffolding SPAT in terms of transfer ability (*adjusted* $p = 0.03^{**}$). Thus, I **observe support for H1b**.

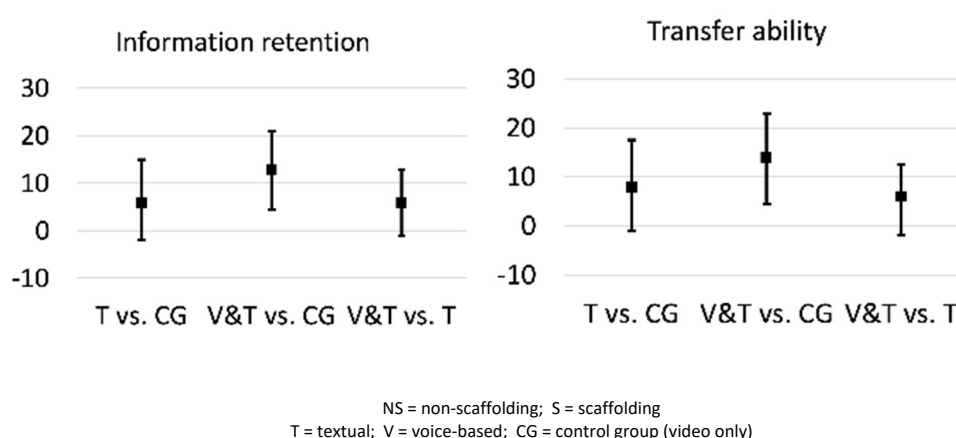


Figure 30. Mean Differences in Scaffolding Manipulation

Source: Own illustration

5.3.9.2 Learner Engagement

In addition to the learning outcome measurements, the main goal of my study was to investigate whether the dynamically scaffolding SPAT Sara is able to increase learner engagement during online video lectures compared to often implemented statically scaffolding SPAT types. To test my hypothesis, I conducted an ANCOVA with learner engagement as the DV and the pretest score as a covariate. Table 31 shows the summarized engagement score including the galvanic skin response and the facial expression score. All scores were standardized between 0 and 1. The engagement score was calculated from the average of galvanic skin response and facial expression.

Measurement	TG1 statically scaffolding SPAT	TG2 dynamically scaffolding SPAT	Control Group
I. Engagement Score	0.31	0.42	0.31
a. Galvanic Skin Response	0.40	0.46	0.33
b. Facial Expression	0.21	0.38	0.28

Table 31. Average Measurement of Galvanic Skin Response and Facial Expression Score

Source: Own illustration

The galvanic skin response shows a higher value for dynamically scaffolding SPAT (0.46) compared to statically scaffolding SPAT (0.40) and the control group (0.33). The facial expression score gives a similar picture. Again, the group with the dynamically scaffolding SPAT shows the highest value (0.38) followed by the control group (0.28) and the statically scaffolding SPAT (0.21).

To check whether there were significant differences between the three groups, I conducted an ANCOVA and continued with a simple linear regression to see whether there were significant differences between statically and dynamically scaffolding types of SPAT. All assumptions for ANCOVA and simple linear regression were met. I ran the ANCOVA with the TG as the independent variable and the pre-test score as a covariate. Results of the test indicate that there are significant differences between the groups ($F(2,360)=46.327$, $p=4.5e-11$, $N=182$). Moreover, the simple regression identifies that there is a significant difference between statically and dynamically scaffolding types of SPAT ($T=7.238$, $p=4.57e-12$, $adjusted\ r^2=0.157$, $N=137$). Thus, I **can accept the hypothesis H2**.

5.3.9.3 Interaction Effect

Regarding the interaction effect for *information retention*, the 2 x 2 ANCOVA revealed a significant effect between the scaffolding mechanism and the modality ($F(1,359)=11.0$, $p = 0.0001^{***}$). Partial eta-squared was 0.03, which is a small effect strength. This result suggests that the effect of a dynamically scaffolding SPAT is stronger when the SPAT is voice-based. Thus, **H3a is supported by the data**. For *transfer ability*, I also saw a significant interaction between non-scaffolding/scaffolding and the voice-based/modality ($p = 0.009^{***}$). Partial eta-squared was 0.019, which is again a small effect. Thus, **H3b is supported by our data**.

5.3.9.4 Qualitative Data

Based on my neurophysiological findings regarding the positive relationship between dynamically scaffolding SPAT and student engagement, I further investigated how students perceived the interactions with dynamically scaffolding SPAT. The three main themes I identified from the open question in the post-survey were *emotional involvement*, *structured thinking*, and *voice usage*. The subtopics and the corresponding frequencies of the responses are presented in Table 32. Below, I briefly discuss the different themes.

	Total	Statically scaffolding SPAT	Dynamically Scaffolding SPAT
	N	Frequency	Frequency
IV. Emotional Involvement			
- Was like talking with a real tutor	19	5	14
- Improves involvement a lot	10	2	8
- Did not feel under pressure compared to dealing with a teacher	4	1	3
V. Structured Thinking			
- Helped me to express the content in my own words	12	1	11
- Helped me to reflect about what was said in the video	11	4	7
- Helped me to improve my understanding of the concepts	5	1	4
VI. Voice Usage			
- Surprisingly accurate	9	4	5
- Exciting to use voice	6	2	4

- Voice was cold and robotic	6	2	4
------------------------------	---	---	---

Table 32. Themes Identified for the Open-ended Question Asking Students About their Experience when Interacting with the SPAT.

Source: Own illustration

Emotional Involvement

This theme relates to the capacity of dynamically scaffolding SPAT to involve students in a human-like interaction. Students in the dynamically scaffolding SPAT group appreciated the use of free utterances when interacting with Sara and that the SPAT was able to adapt its answers accordingly. As one student from the dynamically scaffolding SPAT group said: “Useful, makes me think about what I have just seen. It improves involvement a lot. It was actually better than the video, as listening to the guy and reading the slides at the same time can be annoying.” Other students mentioned that it almost felt like having a conversation with a real tutor. For example, one student mentioned: “Great experience, sometimes Sara was not able to understand me but all in all it felt like chatting with a real tutor.” Another student said: “Chatting with Sara helped me to rethink the different video parts, almost like a human tutor.” For other students, it felt like a kind of relationship. For example, one student said: “It was a human–machine relationship. Not so many emotions were transmitted but she [SPAT Sara] did a good job.” The coding of the themes showed that these kinds of statements about emotional involvement were mentioned much less in the static SPAT group.

Structured Thinking

This theme relates to the ability of students to organize their own thinking structures. Many students mentioned that the dialog with the SPAT helped them to rethink the just learned content from the video segment and use their own words to repeat the content. For example, one student stated: “Both videos helped me to think about what was said by the instructor. Voice-based SPAT did sometimes not catch the right utterance of the student. Though, I preferred voice over text because it is more convenient and helped me to structure my thinking processes.” Some students also mentioned that they preferred the dynamically scaffolding SPAT to the statically scaffolding SPAT in relation to their structuring of thinking processes. For example, one student commented: “I liked writing text instead of clicking ‘next’, because this way I can express the content in my own words.”

Voice Usage

This theme relates to students’ impressions while interacting with the SPAT by voice. While some of the students were enthusiastic about using voice to interact with the SPAT, others were slightly disillusioned about the current state of the technology. The positive aspects students mentioned were that they were surprised about the preciseness of the voice processing and that talking things out

loud helped them to learn more deeply. For example, one student mentioned: “It was fascinating how much it understood and could say if I was correct. It helped me to express my own thoughts.” Another student mentioned: “The video with the voice was better because it forces me to think about the content rather than simply clicking ‘next’”. Others felt slightly annoyed because the current technology seems unable to approach human–human communication. For example, one student commented: “When I could communicate in text it was easier for me to form the answers since I could make corrections, check my answers, and felt like I have less pressure while answering.” Another student mentioned: “Sometimes it didn’t understand me correctly, so I had to adjust the way I was talking.”

5.3.10 Discussion

The aim of this study was to investigate the effect of DP3 (dynamic scaffolding) by using a dynamically scaffolding SPAT incorporated into online video lectures to improve information retention and transfer ability. I proposed that a dynamically scaffolding SPAT that scaffolds students’ understanding during video instruction increases information retention (i.e., helps them remember concepts better) and transfer ability (i.e., helps them apply the acquired knowledge to solve novel problems). The underlying rationale for these hypotheses was that dynamically scaffolding SPATs are able to detect knowledge gaps of students, can react accordingly, and are able to freely interact with them in a similar manner to the gold standard of educator–student interactions.

Hypotheses 1a and 1b confirmed that dynamically scaffolding SPAT can better create a meaningful interaction with students compared to non-scaffolding-based SPAT, resulting in better information retention and transfer ability. Scaffolding-based SPAT might help students to dive deeper into the topic compared to non-scaffolding-based, more traditional SPAT. Furthermore, I found that dynamically-scaffolding SPAT can also increase learner engagement (H2), and voice-based SPAT can further strengthen the scaffolding effect for information retention and transfer ability (Hypotheses 3a and 3b). My study has several implications for the design of future SPAT in online education. I thereby build on previous findings that have investigated beneficial effects of SPAT in online learning and similar environments. For example, Ruan et al. (Ruan et al. 2019) discovered that Quiz-Bot yields increases in factual knowledge compared to more common learning aids, and Lin et al. (Lin et al. 2020) discovered that lessons instructed in a conversational style enhance information retention. Litman et al. (2006) compared voice-based vs. text-based tutoring and found that adding spoken language capabilities increases the performance in human tutoring but not in computer tutoring.

One main reason for the positive relationship between dynamically scaffolding SPAT and student learning outcomes and engagement might be that students are more emotionally involved in the interaction (Theme 1). Dynamically scaffolding SPAT can react to students’ utterances individually, in a similar manner to

human–human communication. This mechanism might help students to become more deeply involved in a conversation, resulting in higher student learning outcomes and engagement. These results match the insights of the ICAP Framework proposed by Chi and Wylie (2014), which states that an interactive, dialoging learning mode is the most effective way of learning. This is also reflected by my qualitative findings, which indicate that the interaction with dynamically scaffolding SPAT almost felt like having a conversation with an educator. When an educator interacts with a student individually, they are able to adapt the answers to the student. Compared to statically scaffolding SPAT, the dynamically scaffolding SPAT might be able to behave more like the gold standard of the scaffolding behavior of an educator. This might be the reason for a student’s deeper involvement in the course materials. The finding is also supported by research in the area of the computers-as-social-actors’ paradigm, which asserts that the more social cues an agent has, the more it feels like talking to a real human (Nass et al. 1994).

Another reason for the positive, quantitative relationship between dynamically scaffolding SPAT and student engagement might be that the individual interaction during video lectures helps students to directly reflect on what they have seen and to express the received information in their own words. This is similar to the results found in Azevedo et al. (2004), who were able to show that students conducted more self-regulating learning activities when receiving dynamic scaffolds. Most of the currently implemented SPAT in online learning relies mainly on standardized, one-size-fits-all hints to improve learning processes, which cannot capture students’ individual utterances (Adamson et al. 2014). For example, Song et al. (2017) developed a SPAT that helps students reflect on their learning on a weekly basis, but which is not able to fully adapt to the students’ answers. Past research in the area of self-talking highlights the important role that articulation plays in the learning process (Lepadatu 2012). For example, Lepadatu et al. (2012) confirmed that students who talked aloud while practicing a ball throwing exercise performed better. This is also shown by my qualitative findings, in which students mentioned that voice input helped them to bring structure to their thinking (theme 2). While some students were surprised at how well the SPAT understood them, the reactions of others clearly showed that voice recognition systems still have a long way to go when it comes to capturing free expressions from users (theme 3). This insight can also be confirmed by past research that deals with voice recognition systems (Grant et al. 2019).

The findings advance the current line of research around SPAT in online learning environments in several ways. First, I can confirm that the findings of past research regarding the importance of dynamically scaffolding mechanisms in SPAT also applies to an online video lecture setting. Scaffolding mechanisms can enable SPAT to identify students’ individual knowledge gaps, react to them, and bring the student back on board again. In the open question section of our post-survey, one of the students said the following: “*Great experience! Sara helped me to think*

about what was said by the instructor and offered help when I was wrong. She helped me to structure my thinking processes." Compared to more traditional SPAT, dynamically scaffolding SPAT does not immediately show the correct answer, but instead tries to help the student build up the missing knowledge and come up with the solution themselves. In the context of online video lectures in particular, it is important that students are able to receive this kind of individual support immediately after not understanding something. Otherwise, they may not be able to process the content that follows, which might result in frustration and less effective learning processes.

Second, based on my findings I recommend equipping future SPAT designs with speech recognition. This is because voice-based SPAT might be able to build a more meaningful interaction with the user than textual methods. Moreover, I was able to detect an interaction effect indicating that dynamically scaffolding is more powerful when it is voice- and text-based. I thereby link this study to research that compares text-based SPAT with voice-based SPAT. In contrast to Litman et al. (2006), I was able to conclude that including voice in computer tutoring dialogs is more effective than text-based only. This extends the results of Novielli et al. (2010), which showed the beneficial effects of voice on social attitude toward SPAT. Moreover, I extend the modality principle (Moreno 2006) by concluding that voice is only beneficial when it is implemented in combination with a dynamically scaffolding mechanism in a SPAT context. This might be because a voice- and scaffolding-based SPAT is better able to imitate a real educator-student interaction. One student said the following: *"Though the SPAT was not able to understand me all the time, it felt a bit like talking to a real person."* My results suggest that it is not enough to use voice-based SPAT in online education. The effect of voice-based systems only comes to bear once the SPAT has implemented an educator-like teaching strategy. For a SPAT to partially imitate an educator, it is essential that the student recognizes the SPAT as a social actor. According to the computers-are-social-actors (CASA) paradigm, people use similar social rules when dealing with computers as with people (Nass et al. 1994). Researchers of the CASA paradigm have assumed that because a computer agent and a human have similar features, the users' social responses are amplified, thus enabling effective interaction with computers (Nass et al. 1997). Embedding a more human-like, voice-based interaction might allow students to perceive a higher level of social presence. This feeling of social presence is usually missing in online learning environments and can influence learning behavior significantly (Richardson and Swan 2003).

Third, our study indicates that publicly available NLP frameworks offer the possibility for SPAT to interact with students freely resulting in a higher quality of interaction. When students need to articulate their answers in a free manner, they need to make more cognitive effort, in a similar manner to when the interaction is with a human educator. This interaction mode can additionally stimulate students' thinking processes, which ultimately leads to better learning outcomes (Im Craik

and Lockhart 1972). As one student put it: “*It was fascinating how much it understood and could say if I was correct.*” However, my research also shows that there is still a lot of room for improvement in NLP and that the quality of interaction becomes higher as the technology advances. Some students were not fully satisfied with the way Sara understood their answers. One student mentioned: “*The system could not always understand my voice, this is annoying.*” With rapid advancements in NLP, future designs of SPAT could use publicly available NLP modules to give students the opportunity to interact freely with the SPAT. Overall, our work provides empirical evidence that using dynamically scaffolding SPAT in online video lectures is beneficial. Nevertheless, introducing it in real online learning environments still requires considerable efforts from educational institutions and lecturers. However, the possibilities for the development and integration of SPAT are becoming easier. For example, it is already possible to develop SPAT without programming knowledge (VoiceFlow 2019). In addition, the integration of SPAT in online video lecture content offers the very promising possibility of creating individual and meaningful interactions in a scalable and personal resource-saving way.

Fourth, I contribute to research in the area of online and blended learning by empirically proving that dynamically scaffolding SPAT built into video lectures can induce a more engaged learning mode in students. A more engaged learning mode is crucial for learning success in online learning, since educators are not able to interact with each student individually.

Fifth, I contribute to NeuroIS research by showing how neurophysiological measurements can be used to understand the impact of conversational systems on learner engagement in online education. To better understand the problem of low engagement and corresponding high attrition rates in online courses, it is crucial to use different measures of learner engagement.

Sixth, I contribute to scaffolding theory by confirming that scaffolding during the instruction phase is only effective when it is dynamic. My findings indicate that static scaffolding produces no different effects to no scaffolding.

In regard to practical implications, this study suggests how SPAT developers and online course providers can design and use SPAT in addition to already existing online video lectures. As indicated by my study, publicly available NLP frameworks such as NLP.js can be powerful tools for creating SPAT that is able to classify intents and to offer dynamic scaffolds.

5.3.11 Limitations and Future research

This study is not without its limitations. First, although our language model almost always recognized a *right*, *wrong*, or *don't know* answer, it made some mistakes, which quickly frustrated students and possibly influenced their learning process. Moreover, the quality of interaction was heavily dependent on our training data

collected from pretests. It would be interesting to see whether SPAT using different NLP frameworks with different training data produces similar results.

Second, interacting with SPAT in a laboratory setting might be different to interacting in real life environments. Students might be more motivated when using the new technology for the first time, which might result in a higher quality of interaction. To partially compensate for that, I asked students about their pre-experience with SPAT, which revealed a rather high experience level. Future research should investigate dynamically scaffolding SPAT in real online courses in which students interact with the SPAT over a whole semester. With the help of qualitative interviews, additional insights can be uncovered.

Third, I tested the SPAT Sara in a narrow context (programming knowledge). I tried to choose a context that is representative of current online courses. Since the number of online programming courses has increased in recent years, I decided to choose such a course. Nevertheless, dynamically scaffolding and voice-based SPAT may achieve other effects in other contexts. Future research should try to ascertain to what extent the subject or domain has an influence on the effectiveness of dynamically scaffolding SPAT in online education.

Fourth, I measured student engagement during a relatively short time period. It is very difficult to make assumptions from these results on how SPAT would have affected the overall student engagement over the full duration of an online course. Nevertheless, I measured student engagement continuously with neurophysiological measures rather than employing a single-point self-reporting of the students. Future research should implement dynamically scaffolding SPAT in online video lectures in real online learning environments in order to measure long-term effects on student engagement.

This study contributes to the question of how to create meaningful interactions in large-scale online courses. The success of SPAT in online learning environments will very much depend on the extent to which educational institutions and lecturers succeed in developing SPAT themselves. Hence, future research should focus on the development of appropriate authoring tools that make it possible to create SPAT without much programming knowledge. These authoring tools should make it easier to design dynamically scaffolding SPAT that includes our proposed DPs. Given that the results from dynamically scaffolding and voice-based SPAT are successful, DP3 (dynamic scaffolding) could also have an effect on areas other than education. For example, SPAT of this kind could also be incorporated into explanatory videos such as video tutorials on using smart systems or voting videos.

5.3.12 Conclusion and Next Steps for the Course of the Dissertation

In this study, I investigated to what extent DP3 (interactive scaffolding) incorporated into a SPAT can solve the challenge of creating meaningful interactions in online video lectures. I implemented a main- and sub-dialog mechanism that led

to a significant increase in students' learner engagement, information retention, and transfer ability. This effect is even stronger in voice-based SPAT systems. These findings from an artificial environment suggest that SPAT can add value to online courses.

As a further step, I wanted to determine whether SPAT can also be beneficial in online learning environments. To this end I integrated SPAT within a real online course unit, to address the second part of RQ2, thereby testing DP4 (engaging dialog flow).

5.4 Using Smart Personal Assistant Technology for Online Learning Activities – Insights from a Field Study

This study complements RQ2b and focuses on DP 4, involving dialog flow. It was performed in a real online learning environment, and helps to answer RQ2b:

RQ2b: How can engaging dialog flows incorporated in Smart Personal Assistant Technology improve students' learning success in online learning environments?

The initial starting point for the answer to this RQ lies in the previously tested DPs and particularly DP4, engaging dialog flow. Based on the initial positive findings of Section 5.3, I created a SPAT for a real online learning environment. The learning content of the course is an introductory unit of management in a higher education setting. To address this RQ2b, I developed and integrated a SPAT in a real online learning environment in which 76 participants had to solve a 20-minute task on the topic “SWOT-Analysis”. One group only received text as the input while the other group was additionally able to interact with SPAT. To implement DP4, the SPAT was equipped with an engaging dialog flow. Both groups had the same amount of time, and the amount of information provided was equivalent for both groups. Based on the ICAP Framework, I argue that the interaction with the SPAT brings participants into a more interactive learning mode, which is considered as the gold-standard of learning.

Properly designed SPAT used in learning activities has the capability to build meaningful interactions with students, bringing them from a passive to an interactive learning mode (Chi and Wylie 2014). I argue that the students have a tendency to exhibit an interactive learning behavior with the help of SPAT, resulting in increased learning outcomes. Chi and Wylie (Chi and Wylie 2014) highlighted exemplary learning activities that help students to move into an interactive learning mode. One of the examples they mentioned was “Asking and answering comprehension questions to a partner”. A SPAT can take the role of such a learning partner. I therefore propose the following hypotheses:

H1a: Students who interact with SPAT during online learning activities achieve higher levels of objective learning outcomes compared to those students who learn without this interaction.

H1b: Students who interact with SPAT during online learning activities achieve a higher level of subjective learning outcome compared to those students who learn without this interaction.

5.4.1 Method

5.4.2 SPAT design

The SPAT was developed based on Python and used the MTurk API to automate the experiment. ParlAI (Miller et al. 2017) was used for higher abstraction and simplification of the MTurk-API in dialogue-oriented applications. Creating an en-

gaging dialog flow isn't a simple process and requires a lot of creativity and planning ahead. When implementing the engaging dialog flow (DP4), I followed a six-step process: 1) decide the SPAT's purpose; 2) create user personas; 3) create a conversation diagram; 4) create conversation scenarios; 5) perform testing; 6) bring it to production.

In step 1, the SPAT should help students to better differentiate the different quadrants of the SWOT Analysis and help them answer the key questions in every quadrant based on a real-life example. In step 2, I created a user persona. The students I addressed with the bot were technologically competent, young, open-minded, goal-oriented, and energetic. In step 3, I created a conversation diagram with different conversational blocks. The conversation diagram consisted of 1) greeting; 2) informing; 3) asking; 4) checking; 5) suggesting; 6) apologizing; and 7) error. With the help of simple boxes and arrows, I mapped out the different dialog flows and corresponding elements and labeled them accordingly. In step 4, based on the structure of step 3, I created conversation scenarios in which I tried to anticipate different student-SPAT dialogs that cover all of my dialog flows from step 3. Figure 31 shows an exemplary dialog with the labeled blocks. In step 5, I created a prototype of the bot and tested it with five students who had similar traits to those of the user persona. In step 6, I used Python to code the bot and make it available to the students via a web interface. I used a proactive and reactive logic where the tutor was able to ask students comprehension questions and was able to respond to students' questions. The guidance and answers given by the tutor and the questions and answers offered by the student were transmitted by chat messages. All five questions can be found in Appendix B.1.

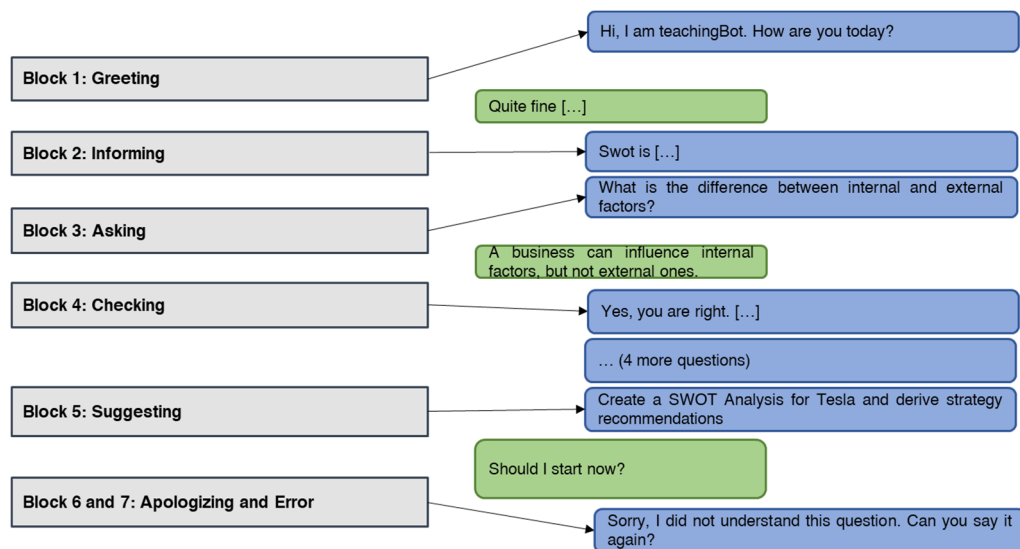


Figure 31. Example Dialog Between SPAT and Student

Source: Own illustration

The SPAT communicates via text and voice on both the input and output sides.

5.4.3 Sample

To answer my research questions and test my hypothesis, I conducted an online experiment with online course students. The tested sample was as follows: 78.9% of TG1 participants ($n = 30$) and 94.7% of the TG2 participants ($n = 36$) lived in the USA; 10.5% of TG1 participants ($n = 4$) and 5.3% of TG2 participants ($n = 2$) were from India; and the remaining four TG1 participants came from the United Kingdom, Brazil, the Netherlands, and Australia. 68.4% of the TG1 participants ($n = 26$) and 63.2% of the TG2 participants ($n = 24$) indicated that they were male. The average age of both TG1 and TG2 was in the category “30 to 34 years old”. Most of the students had at least a high school degree and no experience with SWOT-Analysis ($> 80\%$). The t-test and the chi-square test were used to check whether the two groups were equal regarding age, gender, nationality, and educational background. No differences were found ($p > 0.05$). Furthermore, I used linear regression to check that none of the mentioned characteristics had an effect on the DVs.

5.4.4 Experimental Procedure

I employed a fully randomized online experiment over a period of two months. The experiment lasted 30 minutes. I chose to use a SWOT analysis task for various different reasons. First, the SWOT analysis can be considered as an introductory management topic in an undergraduate online course and is thus representative. Second, I wanted to ensure that the participants did not have much pre-knowledge of the task. Third, my aim was to not only ask participants about reproducing information but also to address higher learning outcomes, and the SWOT Analysis is appropriate for “Apply” tasks.

The experiment consisted of two phases: the instruction phase and the exercise phase. In the instruction phase, students of both groups received a text about the main components of a SWOT analysis and its purpose. The experiment group was able to have an additional conversation with the SPAT. The SPAT asked questions relating to the elements of a SWOT analysis and additional examples. The time given for the instruction phase was equal in both groups, and the information provided by the SPAT was equivalent to the information in the text. The outline of the experiment procedure can be seen in Figure 32.

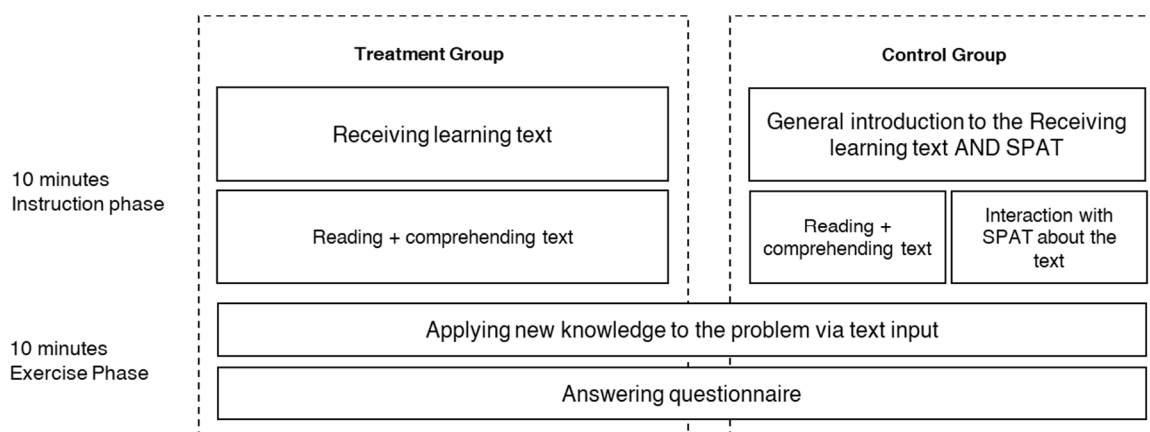


Figure 32. Experimental Procedure

Source: Own illustration

In the exercise phase, the participants in the control and experiment group were supposed to apply their newly learned knowledge by applying a SWOT analysis to the company “Tesla”. The task was as follows:

Create a SWOT Analysis for Tesla and derive strategy recommendations from it.

After the exercise phase, the students answered a questionnaire devised to collect further constructs, and these are described and evaluated in the following sections. The questionnaire was released and answered immediately after the task was completed. The questionnaire included an open question, the purpose of which was to qualitatively examine whether and in what way the interaction with the SPAT was helpful and meaningful. This question was only given to the participants of the experiment group.

5.4.4.1 Measurement and Analysis

The **objective learning outcome** was measured by evaluating the solutions the students proposed for the tasks. The evaluation of the solution was conducted by two raters using a commonly discussed evaluation framework. The two raters rated the solutions blindly and independently of each other. Both raters were experienced in rating SWOT analysis and had been active in university teaching in the past. The final objective learning outcome score was calculated by the average of the two raters’ scores. Each correct example for a category within the SWOT analysis was evaluated with the same number of points.

The **subjective learning outcome** was measured in the post-test questionnaire using a seven point Likert scale ranging from strongly agree (1) to strongly disagree (7). The variable was intended to measure the learning outcome from the student's point of view and thus reflect the construct described in the theory. Before evaluating the data, I inverted the negatively interpreted items so that a higher value in the data analysis corresponded to a better result. In the case of a negative item, this inversion was not necessary. The items are based on the

method devised by Gupta and Bostrom (2013) and, according to Cronbach's Alpha, have an internal consistency of 0.91 and thus exceed the standard of 0.8.

Additionally, I measured the **used time** for the whole exercise to see whether there were differences between the two groups. The time was logged when the task was completed by the participant or the assigned time had expired (maximum 30 minutes). The duration of the questionnaire was not considered, because the focus was on the learning process itself.

In order to examine to what extent the SPAT might have helped students to understand the SWOT analysis, I asked the experiment group the following question:

“Did the dialog with the SPAT support you in learning SWOT analysis and in what way? (1-2 sentences/bullet points).”

Following the method devised by Ryan and Bernard (Ryan and Bernard 2003), I identified topics from the answers by using “repetitions” and “similarities & differences”. For data analysis, I used a linear regression model to see whether the scores were significantly different. I used R for the statistical evaluation. In the further data analysis, it was first examined whether the assumptions of linear regression—in particular, normal distribution, homoscedasticity, and linearity—are given for the established models. All assumptions were met. The R library “gvlma” created by Peña and Slate (Peña and Slate 2006) was used for this test. Moreover, I calculated the effect strength according to Cohen’s d (Cohen 1992).

5.4.4.2 Results

Table 2 shows the averages and SDs for the data collected for all variables. Brackets next to the variable name indicate the ranges that the values could assume. A higher average corresponds to a better result.

	Control Group		Experiment Group	
	Mean	SD	Mean	SD
Objective learning outcome (0-12)	6,1578	3,7885	7,7631	2,7453
Subjective learning outcome (1-7)	4,0368	0,9697	4,4316	1,0214
Used time in min. (0-∞)	18.7	208	19.3	223

Table 33. Mean and Standard Deviation for the Variables

Source: Own illustration

The results of the regressions for hypotheses 1a and 1b are presented in Table 3. The coefficients marked with * declare significance levels as follows: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$.

I identified a significant difference in objective learning outcome ($t = 2.115$, $p = 0.0378$). Cohen's d is 0.4852. Thus, I can support H1a. This is almost a medium effect according to the rule of thumb of Cohen. For the subjective learning outcome (H1b), no statistically significant difference could be detected ($t = 1.728$, $p = 0.0882$). If the weaker standard of $p = 0.1$ is applied, there is a statistically significant difference. Cohen's d would then indicate a small effect with $d = 0.396$. Moreover, no significant difference in the time used could be identified ($t = 0.096$, $p = 0.924$). These data are summarized in Table 34.

Variable	Coefficient	p-value
H1a: Objective learning outcome	1,6053	0,0378**
H1b: Subjective learning outcome	0,3947	0,0882*

Table 34. Regression Results

Source: Own illustration

The qualitative data reveals that a total of 76.3% ($n = 29$) of the participants responded positively to the questions. 10.5% ($n = 4$) of the participants tended to express a negative opinion of the questions. The remaining 13.2% ($n = 5$) of the answers either reported a neutral statement or could not be assigned because they did not directly refer to the question. In total, I was able to identify three major topics. These are displayed below and describe how the students were supported by the questions.

Practical relevance and examples: this denotes that the SPAT helped students transfer the acquired knowledge to real-world examples. The students were able to more easily connect the learning content with their existing knowledge experiences.

Structuring aid: this refers to the ability of the SPAT to help participants structure the learning content. In this way, it was easier for them to remember the content when conducting the task. The participants emphasized that this structuring aid in particular enabled them to reduce the SWOT analysis to its essential content.

Internalization: this refers less to the understanding of learning than to the fact that the SPAT makes the learning process more memorable and challenges the student interactively. The direct method of conversation helps students to internalize the learning content in its essential parts so that it can be retrieved more easily later.

5.4.4.3 Discussion

In this study, I investigated whether DP4 (engaging dialog flows) incorporated into SPAT are able to increase learning outcomes in a real-life online course learning environment. Drawing on the concept of the ICAP framework and a five-step process, I developed an engaging dialog flow and shared it with the students via a web interface. The findings revealed that participants interacting with SPAT were able to significantly increase their learning outcomes compared to participants who had no interaction. This is in line with past research in the field of multimedia learning that was able to show that interactivity is a key DP when designing learning materials (Evans and Gibbons 2007). For example, Evans and Gibbons (2007) discovered that students using an interactive learning system outperformed those using a non-interactive system. Moreover, the results of my study confirm the positive effects revealed in existing research on ITS (Kulik and Fletcher 2016). Regarding the subjective learning outcome, no significant differences were found. One reason for this might be that the conversation with the SPAT enabled students to reflect on their knowledge gaps resulting in a less well perceived learning effect.

It is noteworthy that although one would expect the participants to need more time to read and interact with the SPAT compared to the time required for traditional learning materials, they did not need additional time. This suggests that the learning process was fundamentally different for the two groups and that the interaction with the SPAT was not just an add-on to the reading of the text. Some participants might have omitted the reading of the text and started to interact with the SPAT earlier, which shows that they preferred the conversational style of interaction.

My work adds to the already shown positive effects of SPAT on learning outcomes. Specifically, this study makes several contributions to two areas of research. First, I contribute to ITS and online learning research by showing that new emerging SPAT is a promising solution for addressing the problem of individual support in large-scale online learning environments such as MOOCs. When students are able to interact with SPAT individually (e.g., ask questions and receive immediate, personalized answers), they might be more motivated, might not abandon the course as quickly, and might show better results. This would help educational institutions and lecturers to reduce costs while improving the quality of their learning material. Second, I contribute to research around the ICAP framework by indicating that IT might be able to bring students into an interactive learning mode, which can be seen as the gold-standard in learning. Past research has mainly focused on human partners endeavoring to boost students' interactive behavior. I show that new emerging types of IT can bridge the gap between students' constructive and interactive learning modes.

This work also has several implications for practice. First, the implementation of SPAT in online learning environments might enrich students' learning outcomes.

SPAT providers such as Amazon or Google offer large ecosystems that allow an educator to create their own SPAT with little or no programming knowledge. Based on the findings of this study, an educator might be motivated to create a SPAT for their own learning environment. Second, MOOC providers such as Coursera and EdX might be inspired to offer SPAT templates to their course providers in order to increase learning outcomes and their own business value.

5.4.4.4 Limitations and Future Research

First, the experiment duration was relatively short and was conducted under positive conditions. Such studies tend to produce better outcomes due to novelty effects and hyperattention to experiment details, such as the interaction with the SPAT (Cheung and Slavin 2013). For future research, it would be helpful to see whether SPAT is able to increase learning outcomes over a longer period of time (e.g., over the period of an online course). Second, I only looked at short-term learning outcomes. To identify the true impact of SPAT on learning would require insights into its effect on long-term learning. Third, I tested the use of the SPAT in a very narrow context (SWOT analysis task). Future research might test SPAT interactions for different kinds of tasks and different kinds of knowledge (factual knowledge, applying knowledge, etc.). Fourth, the development of such a SPAT is time-consuming and is not automatically transferable to different learning contexts. This raises the question of the practical value of SPAT. Future research should try to develop artifacts, DPs, or patterns that help educators to create their own SPAT in an efficient way.

Conclusion and Next Steps for the Course of the Dissertation

This work provides the first evidence that SPAT with engaging dialog flows (DP4) can increase learning outcomes in a real online learning environment. I designed and incorporated an engaging dialog flow into a SPAT where participants had to solve a SWOT analysis task. One group received text material only, while the other group received text material and was able to interact with the SPAT. The SPAT interactions helped the students to engage with the learning material in a meaningful way, resulting in increased learning outcomes. These findings suggest that SPAT interactions in online learning activities might be able to induce an interactive learning behavior in students, which is considered as the gold standard according to the ICAP-framework (Chi and Wylie 2014). These findings contribute to ITS and online learning research. While further research is required, it appears that the use of new emerging SPAT is a promising enrichment for online learning environments.

After answering RQ2, I was able to prove that the implementation of SPAT with the proposed DPs 1–4 is promising for online and blended learning environments. However, simply demonstrating the benefits does not mean that educators at different levels and in different environments will use the technology. Despite the shown benefits, the wide scale adoption of SPAT in these learning environments

remains a big issue. To address this challenge, the secret lies in enabling teachers to design and create SPAT themselves, so that they are independent of software designers and can use these systems flexibly and confidently in their own way. The third RQ is therefore devoted to the question of how educators themselves can be empowered to use SPAT in their own teaching and learning environments.

6 Empowering Educators to Design and Embed Smart Personal Assistant Technology in their own Learning Environments³

In this section of my dissertation, I address the third RQ and the fifth DP, “accessibility”, that deals with how educators can be empowered to create their own SPAT in online and blended learning environments that positively impact students’ learning processes and the resulting learning outcomes. The RQ is stated as follows:

How can non-tech-savvy educators design and embed Smart Personal Assistants in their own online and blended learning environments?

I answer this RQ via two individual studies. In a first step, I develop a process model that guides educators in designing and creating SPAT in their own learning environment. Based on that process, the second step involves the development of a SPAT platform implemented as a web-based single page application that should help educators to follow the steps of the process model of RQ3a.

6.1 Process Model that Helps Educators to Create and Embed their Own Smart Personal Assistants

6.1.1 Introduction

In this section of my dissertation, I will address DP5 (accessibility) and RQ3a, which deals with how non-tech-savvy educators can create their own SPAT in online and blended learning environments. This study answers RQ3a:

What are the relevant steps that help non-tech-savvy educators to create SPAT for their own learning environments?

This study is based on RQ2, the answering of which provided evidence that SPAT can be beneficial for online and blended learning environments. Based on this knowledge, I analyzed DP5 (accessibility) and accompanied educators across different levels to find out how they would use a SPAT platform.

Past research in the field of ITS has mainly focused on SPAT design knowledge for software designers (Barbhuiya et al. 2011). Moreover, recent publications regarding SPAT in learning environments has focused on specific implementations and has not provided transferable insights and in-depth evaluations (Arend 2018). For example, Arend (2018) developed a specific SPAT with the help of Apple’s Siri in a secondary school class for a specific homework assignment without comparing the learning success with a control group. Little is known about how educators can design effective SPAT for their own learning environments (Hobert and Wolff 2019). Hence, the objective of this study is to create a method that allows researchers and educators to design effective SPAT as learning tutors

³ The insights presented in this section are partly based on various papers (Winkler et al. (2020a); Winkler and Söllner (2020)). I thank my collaborators and the anonymous reviewers and attendees of the HICSS 2020 and ECIS 2020 for their valuable feedback on my work.

to address the problem of individual support in large-scale learning environments. I thereby concur with Hobert and Wolff (2019) and Kim and Baylor (2016) in stating that there is a need for generalized design knowledge for SPAT in education.

My design science research project is grounded on a constructivist view of learning. Specifically, I use the ICAP-framework proposed by Chi and Wylie (2014) as my kernel theory. To answer my research questions, I present three cycles of a design science research (DSR) approach as described by Hevner (2007). In the *first cycle*, I derive requirements for the process model from literature and conduct qualitative interviews with educators and students. Based on this, I formulate the first version of the process model and evaluate it with the help of a focus group. In the *second cycle*, I explain the process model to a selection of educators across different levels and let them use the process model to create their own prototype-based SPAT. The process model should help educators to create SPAT that helps them to offer individual support in online and blended learning environments and helps to bring students into an interactive learning mode, a behavior that is regarded as the gold standard according to the ICAP-framework. In the next subsection, I present the existing research concerning end-user development, which forms the theoretical basis for my process model and prototype design.

6.1.2 End User Development

Over the past 20 years, researchers have studied several different approaches that empower users to create their own software programs in order to solve their problems. End User Development (EUD) aims at enabling users to participate in the design and development of their own software artifacts, not only during the design period, but also during their actual use. As a result of the many possibilities provided by recent technology, such as the Internet of Things (IoT) and smart devices, EUD creates methods, situations, and socio-technical environments that allow and empower end users to express themselves and remain independent of high-tech scribes (Fischer et al. 2017) in crafting both software artifacts and hardware technology (e.g., smart objects). EUD is inherently different to traditional software development, and it is often not enough to support EUD by simply mimicking traditional software engineering approaches. Indeed, although end users are usually expert in some particular domain, they tend not to be trained in programming languages and development processes, and nor are they interested in learning them. Consequently, to support EUD activities, end users need to be provided with appropriate tools, communication infrastructures, and development frameworks that are easy to use and to learn, and that are also easily integrated into the domain practice. Previous researchers have used a variety of techniques to help users create their own artifacts. The most prominent method that Barricelli et al. (2019) discovered was a component-based approach (48% of all papers in their review). The component-based approach encompasses composition of digital artifacts by means of visual programming environments based on domain-

specific concepts represented either as 2D or 3D objects. The composition activity is usually performed through direct manipulation, and the interaction metaphor may vary from jigsaw puzzle to boxes and wires, and from pipes to 3D blocks.

6.1.3 Research Methodology

I followed a design science research (DSR) approach to answer the proposed research question. Specifically, I relied on Hevner's three cycle view to structure the research process (see Figure 33). This approach is well suited to my project on account of its three-cycle view. The relevance cycle connects the application domain of the research project with the design science activities. The rigor cycle ensures that the design science activities consider the existing knowledge base of scientific foundations, experience, and expertise. The central design cycle iterates between the core activities of building and evaluating the design artifact (Hevner 2007).

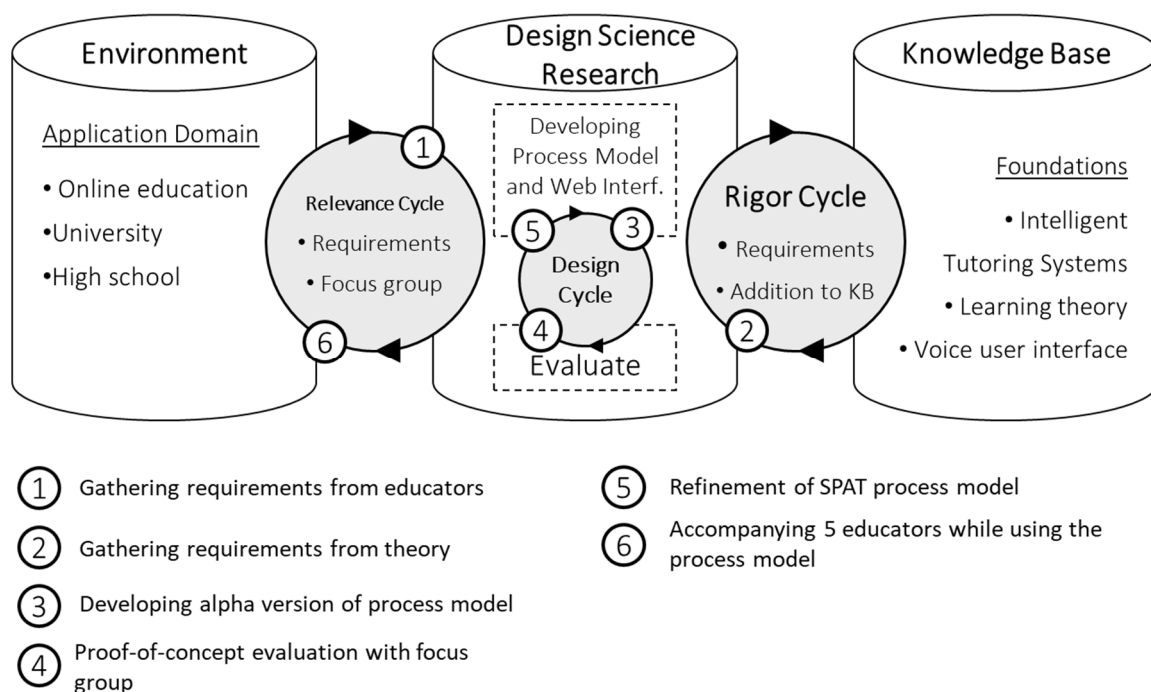


Figure 33. Design Science Research Process

Source: Own illustration

In step 1, I initiate the *relevance cycle* by gathering requirements from educator and student interviews. In step 2, I initiate the *rigor cycle*, gathering requirements from theory using three different perspectives. In step 3, I initiate the *design cycle* by developing the alpha version of the process model. In step 4, I conduct a first proof-of-concept evaluation with the help of a student focus group discussion. This evaluation helped to test my first design hypothesis: *The process model is logically derived from the requirements*. In steps 5 and 6, I build on these findings and refine and explain them to educators. In this manner, I test my second design

hypothesis: *The process model is useful for practitioners*. In the final step, I discuss the results and create an outlook for the next study

6.1.4 Designing the Process Model

6.1.4.1 Step 1: Gathering Requirements from Students and Educators

My application domain involves students and educators from different levels and teaching formats (educators from online and blended learning environments, and students from university, high school, and secondary school). The main goal of my method is for mainstream educators, after a short training period learning the process model and its description, to be able to create prototype-based SPAT for their own learning environments. Prototype means that they can create a basic interaction with the help of a chatbot prototyping tool that is currently on the market (e.g., VoiceFlow). As a first step, I wanted to achieve a basic understanding of the current ways that educators implement these systems, the nature of students' learning preferences and habits in their everyday life, and what students think of using SPAT to enrich existing learning environments. Hence, I conducted 20 interviews with students from high schools and universities. Moreover, I wanted to gather a basic understanding of educators' needs when developing and embedding a SPAT in their learning environments. Thus, I conduct four interviews with educators. Table 35 lists the characteristics of the interview partners. All the interviews were semi-structured and lasted approximately 30 minutes each. I used a paper-based mock-up of a SPAT dialogue for students and a mock-up of the process model for educators.

Interviewees	Characteristics
Students (N=20)	Average Age: 21.35; level of education: 5 x high school, 10 x bachelor, 5 x master; nationality: 8 x Swiss, 11 x German, 1 x Italian; gender: 8 women, 12 men
Educators (N=4)	Teaching experience: 5 to 25 years; type of school: 1 x high school teacher, 2 x bachelor's lecturer, 1 x master's lecturer; experience with different learning environments: 1 x online course, 1 x mass lecture, 2 x small classes

Table 35. Sample Characteristics

Source: Own illustration

I created the interview questions in an internal workshop with the research team. In the first part of the student and educator interviews, I asked: What are the problems you encounter in a course in regard to individual support? Can you explain it based on concrete examples? In the second part, I showed students a

mock-up on a piece of paper demonstrating a student–SPAT learning dialogue. In contrast, I showed educators a first draft of the process model. Based on this, I asked students about possible requirements for using SPAT and asked educators about the requirements of the process model itself. I transcribed the interviews and analyzed them using the method of user stories (US) proposed by Cohn (2004). I identified USs in the interviews, coded, clustered, and translated the USs into requirements for the process model. In the interviews, students often complained about a lack of individual support during the learning phases. Educators often complained that they would need too much time to deploy a technology solution for their learning environment. See Table 36 for more detailed USs and requirements.

User Stories Students (USS)	Requirements from Practice (RP)
USS1: As a <i>student</i> , I want the SPAT to give me immediate and detailed feedback on my responses.	RP1: The process model should help educators to include an evaluation mechanism for students' utterances.
USS2: As a <i>student</i> , I want the SPAT to reply as fast as a human in a conversation.	RP2: The process model should help educators in choosing a SPAT platform with a fast hardware and software architecture.
USS3: As a <i>student</i> , I want to be able to use the SPAT any time, any place.	RP3: The process model should help educators to build SPAT that is available any time, any place.
USS4: As a <i>student</i> , I want to access the SPAT via voice similar to smartphone assistants (e.g., Google's Assistant or Apple's Siri).	RP4: The process model should help educators to build voice-based SPAT.
User Stories Educators (USE)	
USE1: As an <i>educator</i> , I want to first analyze the learning goals, contents, and methods in which I want to use the SPAT.	RP5: The process model should include a step where educators can analyze the intended use scenario.
USE2: As an <i>educator</i> , I want to design SPAT by myself without needing a lot of technological knowledge.	RP6: The process model should be aimed at non-tech savvy educators and should help them with easy-to-use tutorials and blueprints that allow

	educators to develop SPAT on their own.
USE3: As an <i>educator</i> , I want it to be possible for the SPAT to be used on different devices (e.g., smartphones, tablets, laptops).	RP7: The process model should help educators to create SPAT that is accessible via different devices.
USE4: As an <i>educator</i> , I first want to test the SPAT by myself before I use it in my classes.	RP8: The process model should include a step where educators can iteratively test and adapt their initial version of a SPAT.

Table 36. User-Stories and Requirements from Students and Educators

Source: Own illustration

6.1.4.2 Step 2: Gathering Requirements from Theory

I initiated the *rigor cycle* by gathering requirements from theory. I conducted a systematic literature review following established methodical approaches established by Cooper (1988) and vom Brocke et al. (2015). Based on this, I performed four steps: 1) I defined the review scope; (2) I conceptualized the topic; (3) I searched the literature; and (4) I analyzed the findings regarding requirements. Regarding step 1 (defining the review scope), I primarily focused the literature review on research outcomes that show successfully implemented SPAT in learning environments. Furthermore, my goal was to identify requirements on a conceptual level with a focus on a representative coverage (Cooper 1988). Regarding step 2 (conceptualizing the topic), I identified three perspectives useful for deriving requirements for the process model: a technical, educational, and voice user interface perspective. I used these three perspectives because learning with a tutor is a very complex phenomenon that is being investigated through different lenses by psychologists, educationists, computer scientists, and various other experts (Lajoie and Azevedo 2006). Regarding step 3 (searching the literature), I conducted a keyword search in databases to identify relevant publications. I started with all eight journals of the AIS Senior Scholars' Basket of Journals as well as the three most prestigious journals in educational research and then widened the range to include other peer-reviewed journals and conferences to cover the three different perspectives. I selected the following databases: "AIS Electronic Library", "ACM Digital Library", "IEEE Xplore Digital Library", "Science Direct", "EBSCOhost Business Source Complete", and "ERIC". Additionally, I used the search terms "smart personal assistant", "learning conversational agent", "voice assistant", "artificial intelligence teaching assistant", "conversational agent", and "method". In total, I obtained 478 articles. I defined criteria for inclusion and exclusion and reviewed titles and abstracts of all search results as a first

step. I only included papers that systematically displayed the development process of the SPAT and similar computer tutors in order to ensure that I derived proper requirements for the process model. Based on this, I selected 36 papers. Regarding step 4 (analysis of the literature), I scanned the 36 papers to determine the steps necessary for developing SPAT and clustered similar requirements, which resulted in four requirement clusters. Table 37 shows the perspectives, an exemplary paper, and the requirement clusters from theory (RT).

Perspectives	Exemplary Paper	Requirements from Theory (RT)
Educational	ICAP-Framework (Chi and Wylie 2014)	RT1: The process model should help educators to include a proactive and reactive interaction logic, where the SPAT and the student both make a contribution.
	Scaffolding (Hannafin et al. 2004)	RT2: The process model should help educators to include different kinds of scaffolds when creating the interaction model.
Technological	Error handling (Guo and Goh 2015)	RT3: The process model should help educators to include error handling mechanisms to resolve errors and dysfunctions so that the user can continue.
Voice User Interface	Engaging Response Structure (Pearl 2016)	RT4: The process model should help educators to use an engaging response structure that is different to visual interfaces.

Table 37. Requirements from Theory

Source: Own illustration

6.1.4.3 Step 3: Developing an Alpha Version of the Process Model based on Requirements

Based on the theoretical background on end user development and the identified requirements from practice and theory, I developed the initial version of the process model. The main goal of the process model is to guide educators in creating

SPAT for their own learning environments without the requirement for great technological knowledge and with a short training period to learn about the method and its description. The development of the process model is aligned to the definition of a method proposed by Brinkkemper (1996): “A method is an approach to perform a systems development project, based on a specific way of thinking, consisting of directions and rules, structured in a systematic way in development activities with corresponding development products.” When designing the process model, I tried to order the requirements in a way that made sense. I used grey circles for representing development activities and grey rectangles for development products. Figure 34 shows the process model and the six steps that guide educators to create their own SPAT and that guide me in creating a web user interface as the next phase of the experiment. In step 1, the educators need to analyze their learning environment and decide which learning goals, content, and method should involve the SPAT. This results in a determination of the use environment. For example, an educator might want to use the SPAT for helping students to repeat English vocabulary in preparation for an exam, or they may want to use the SPAT for introducing a new topic.

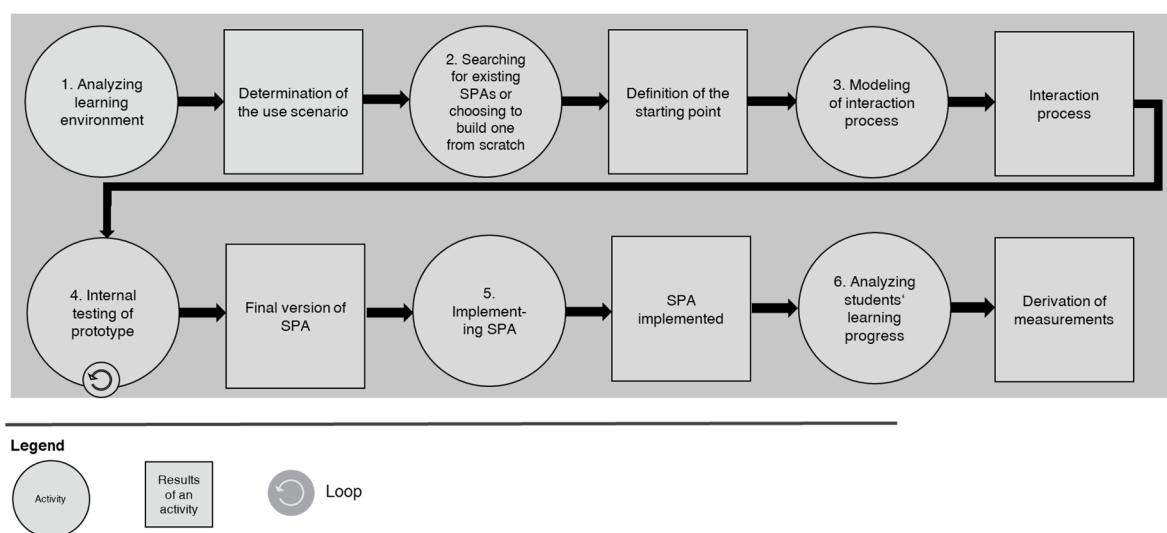


Figure 34. SPAT Process Model for Educators

Source: Own illustration

In step 2, educators can search for already existing SPAT in a database and possibly adapt it, which marks their starting point. In step 3, the educator needs to design the interaction flow by creating a conversation diagram and sample dialogs between the SPAT and the student. The educators model the desired interaction process between the student and the SPAT and add possible students' utterances and SPAT responses at each point of interaction. For modeling the interaction process, educators can use prototyping tools or simply a piece of paper on which they write down the statements the SPAT should make and make a note of what they anticipate the students will say. For example, educators could specify which kind of vocabulary the SPAT should use with the students and

which kind of students' answers might be wrong or right. The educator should then integrate the interaction process on a SPAT platform (such as AVS or Google's Dialogflow). This integration process very much depends on the platform type. Generally, educators need to add possible students' intents and set the rules for the intents. The easiest way for the educators to achieve this might be to employ a drag and drop solution. For example, they can state that when a student says X, the SPAT should say whether the answer is right or wrong and, depending on the student's answer, continue with a different path. In step 4, after creating the interaction model, the educator starts to test the initial prototype internally or externally with a few students in order to refine the interaction flow. In step 5, after some loops of adapting and testing, the SPAT can be shared with the students and implemented in the learning environment. The students should be able to easily access the SPAT via smartphone. In step 6, students and educators alike should be able to analyze the learning progress and might derive measurements from it.

6.1.4.4 Step 4: Proof-of-Concept Evaluation and Refinement of the Process Model

According to Sonnenberg and vom Brocke (2012), for evaluations it is important to focus on two aspects: (1) the constituents of the artifact and the design decisions taken; and (2) the evaluation of the usefulness of the artifact. My first evaluation thus concentrates on the constituents of the artifact. Focus group discussions are considered a suitable method for evaluating designs within DSR projects (Sonnenberg and Vom Brocke 2012; Venable and Baskerville 2012). The goal of the focus group discussion was to check whether the translation of the requirements into the process model made sense. I thereby tested my first design hypothesis: *The process model is logically derived from the requirements*. Moreover, I wanted to gather further CRs for improving the process model. The focus group discussion lasted 60 minutes and consisted of six master's and four undergraduate students with an average age of 23.5 ($m=3$, $w=7$) who were sourced from a Swiss management university (involving different students to those used in step 1). The focus group discussion was structured as follows. In the first part, one of the researchers introduced the goal of the group discussion. In the second part, I explained to students the derived requirements and the corresponding process model. I asked students to identify possible, logical breaks and to collect further CRs. For example, the students suggested adding the loop sign to the prototype testing phase (step 4) to show that this step is iterative rather than a waterfall type of process. Moreover, they mentioned that step 3, "creating the interaction flow", is critical and should be described carefully and supported with templates and examples so that the process model can clearly be seen to benefit the educator.

6.1.4.5 Step 5 and 6: Refinement of the Process Model and Proof-of-Usefulness Evaluation with Educators

After the focus group evaluation, I refined the process model and tested the next design hypothesis: *The process is useful for practitioners*. To test this hypothesis, over a period of four weeks I accompanied four educators who teach in online courses at different levels (2 x public university, high school, primary school, with different educators to those chosen for the interviews). The educators received a one-hour introduction to the process model and also received instructional material to guide them through the steps. Next, they had to create a prototype-based SPAT for their own online or blended learning environment over the period of four weeks and kept a journal recording their experiences while using the process model. Furthermore, I held semi-structured interviews with the educators after the four weeks. Based on the interview data and the journals, I identified pains (P) and CRs according to the method of Ryan and Bernard (2003). Specifically, I used the keywords-in-context method for this study. With the help of this technique, I identified key words indicating pains and CRs when using SPAT. I also searched the corpus of the text to find all instances of the word or phrase. Each time I found a word, I made a note of it and its immediate context. I identified Ps and CRs by physically sorting examples with a similar meaning into categories (Ryan and Bernard 2003). Table 38 displays the identified Ps and the CRs for further improving the process model. Again, I used USs by Cohn (2004).

Identified Pains (by Lecturers)	Corresponding CRs
P1: As an educator, it was hard for me to go from one step to the other. The steps in themselves are clear, but I don't know how to switch to the next step.	CR1: The process model should more clearly guide the user in going from one step to the other.
P2: As an educator, as soon as it became technical, I was overwhelmed and was not sure how to conduct the step. Step 3 was particularly hard to manage without technical knowledge.	CR2: The process model should use interfaces such as a wizard, where educators can follow step by step to create a SPAT, especially in step 3.
P3: As an educator, I was not sure how to create an interaction model from scratch.	CR3: The process model should offer a drag-and-drop interface where educators can create an interaction model based on example entries.
P4: As an educator, I have privacy concerns regarding the students'	CR4: The process model should tell the user what happens to their data and

data. I don't want to give big players such as Amazon or Google sensible data about my students' learning difficulties and progress. I want to see what happened to the data.	should keep personal data within their own database. Moreover, an educator should be able to delete their data.
---	---

Table 38. Ps and Corresponding CRs

Source: Own illustration

Overall, the main points I identified from the analyzed transcripts and journals were that the process model and its corresponding key questions are still not attuned enough to the needs of educators, particularly in regard to the wish to jump from one step to the next (P1). Educators wished to see more clearly how they can use the output of one step as input for the next step (CR1). Moreover, they struggled when they had to technically transfer the interaction flow into a SPAT platform in step 3 (P2) and when they had to create an interaction model from scratch (P3). The process model should provide interactive, wizard-like help to conduct the technical steps (CR2) and to offer pre-defined templates with examples that enable educators to intuitively see how it works without the need for in-depth technical knowledge (CR3). Furthermore, educators do not want to share students' data with big players such as Amazon and Google and want to have control over their private data (P4). The process model should therefore tell the user what happens to their data or keep the data in the educator's own database and should provide the lecturer with the option to delete the data after creating the SPAT (CR4).

6.1.5 Discussion

In this study, I developed a process model that helps educators create SPAT for their own learning environments. This study forms the first part of the answer to RQ3. New emerging SPAT such as Amazon's Alexa and Google's Assistant promises a helpful enrichment of learning environments by offering individual support that would otherwise not be possible due to financial and organizational restrictions. SPAT can help educators to address the problem of individual support in large-scale learning environments. I proposed two main design hypotheses to test whether the process model was able to fulfill the overarching goal. Within my design hypothesis 1 (*The process model is logically derived from the requirements*), I evaluated whether I had correctly shaped the method according to students' and educators' requirements. In design hypothesis 2, (*The process model is useful for practitioners*), I explained the process model to educators and asked them to judge whether this method helps them to create their own SPAT.

In the remainder of this section, I describe this study's contributions to research and practice, as well as its limitations, and make the transition to the second study that answers RQ3. My work makes several contributions to research in this field. First, it contributes to research regarding the ICAP-framework by showing that

educators find it useful in bringing students into an interactive learning behavior considered as the gold standard. When students interact with SPAT, they dive deeper into the learning material, which results in an increased learning outcome. Second, it contributes to ITS research by proposing a new process model that helps educators to use new emerging types of SPAT as learning tutors. Until now, to the best of my knowledge, there has been no such method for educators. Existing methods regarding the design of ITS have addressed software designers. Consequently, there was a gap between the technological know-how of software designers and the pedagogical know-how of educators. As a result, ITS were not very widespread in learning environments. Third, this method helps researchers to create new emerging SPAT for learning environments in order to further investigate their value compared to other kinds of learning aids.

My work also has several implications for practice. First, the proposed process model provides initial help to educators who want to offer individual support to students without increasing costs or time resources. Since there has been a significant increase in online and distributed learning environments in recent years, SPAT as a new type of ITS is more relevant than ever before. Second, our method helps educators to transfer SPAT from one context to another without being dependent on the technological know-how of software designers.

6.1.6 Limitations and Future Outlook

A number of limitations have to be considered with respect to this study. First, the list of requirements is derived from specific theoretical perspectives and from a certain selection of students and educators. It is likely that the requirements would be different if I had used different perspectives and interview partners. However, I tried to include the most important theoretical perspectives and select a representative sample of interview partners. Additionally, the educators all used Amazon's Alexa platform to create their SPAT prototypes. Other platforms might have led to different results.

6.1.7 Conclusion and Next Steps for the Course of the Dissertation

In this study, I proposed and evaluated a process model that should help educators to create their own SPAT. The process starts with analyzing the desired learning environment and follows with the creation and testing of the interaction model before the educator can implement it into their learning environment. However, a process model alone is not enough to empower educators to use the full potential of SPAT for their own learning environments. To enable the widespread use of SPAT in learning environments and thereby address RQ3, the final study in this dissertation creates a SPAT platform based on the process model that helps educators to design and deploy their own SPAT.

6.2 SPAT Platform to Empower Educators to Create and Embed Smart Personal Assistants in their own Learning Environments

6.2.1 Introduction

In this section of my dissertation I address RQ3b, investigating how educators can include technology in their learning environments. The research question RQ3b is as follows:

How can non-tech-savvy educators be empowered to design and embed SPAT in their own learning environments?

The developed web user interface in this study is based on the process model Section 6.1 described, where I tested how educators proceed when implementing SPAT in their own learning environments. Based on this knowledge, I created a SPAT platform where educators can create and embed SPAT on their own.

There is consensus in past literature that SPAT is able to successfully meet the individual needs of students and increase their learning outcomes (Kulik and Fletcher 2016). Despite these proven effects, the current distribution of SPAT in online and blended learning environments is scarce. If used at all, SPAT is implemented in tech-savvy online courses, where educators have the technological know-how to create a SPAT from scratch (Hooshyar et al. 2016). Little is known about how non-tech-savvy educators can design effective SPAT for their own learning environments (Sottolare et al. 2016). Hence, the goal of this study is to create a SPAT platform that allows non-tech-savvy educators to create SPAT that offers individual support. In the first step, I visualized the user interface for each step of the process model with the help of a prototyping design tool. I then conducted a first design evaluation and technically implemented it with the help of the JavaScript framework React, the CSS framework MaterialUI, Express, Nodejs, and SQL. I then evaluated the alpha version of the user interface, called myteachingbot.com, with 38 educators and created a beta version of the platform. The results show that the SPAT platform is able to empower educators to design their own SPAT to offer individual scalable support. In this way, SPAT can enter more online and blended learning environments in the future. Particularly in online learning, SPAT can help students to receive individual support and can thus decrease attrition rates.

6.2.2 Theoretical Background

6.2.2.1 Existing Authoring Tools for Developing ITS

Sottolare et al. (2016) dedicated the third book in their series on ITS to authoring tools. In the book, they provided five examples of successful authoring tools currently on the market:

- Authoring Software Platform for Intelligent Resource in Education (ASPIRE) created by the Intelligent Computer Tutoring Group at the University

of Canterbury in New Zealand. The platform enables experts to create constraint-based tutors through the generation of domain model supplemental information via interaction with the system (Sottolare et al. 2016).

- AutoTutor Authoring Tools developed by the University of Memphis. These empower users to create conversational scripts via a web-based interface and have recently been upgraded to simplify the authoring process to a level that helps students to create ITS for themselves (Nye et al. 2014).
- The Cognitive Tutor Authoring Tools (CTAT), created at Mellon University. CTAT allows authors to demonstrate model solutions rapidly. In general, it supports tutors who work with individualized, step-by-step guidance during complex problem-solving.
- The GIFT Authoring Tools, developed by the Army Research Labs (ARL). The source code for the tools is open-sourced. GIFT was created to realize the US Army Learning Model self-regulated learning capability and to reduce the time, cost and skills needed to author intelligent tutoring systems. The aim of GIFT is to integrate pedagogic practices in GIFT-generated ITS, and it is able to extract models in systems that are not generated by GIFT. With a target group of students, the system offers automatic and blended instructions where GIFTs can be used to support curriculum objectives (*Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym3)*, 2015).
- The Situated Learning authoring tool (SitPed), developed by the University of Southern California. This tool specializes in preview-based authoring and enables a non-tech-savvy author to simulate the experience of a student while simultaneously adapting the conversation script (Chad Lane et al. 2015).

The difficulty of creating an authoring tool that non-tech-savvy educators find easy to use remains one of the major obstacles to successful development and implementation. Sottolare (2016) lists the four main barriers that prevent non-tech-savvy users adopting ITS:

- Specialized skills needed (e.g., computer programming)
- Too much time and cost effort to create an ITS
- Too much time to retrieve and organize authoring content
- No standardized norms for authoring an ITS

My research question addresses the first three of these barriers.

6.2.3 Technical Development of MyTeachingBot

In this section, I present the technical development of MyTeachingBot. The technical development follows the seven steps depicted in Figure 35.

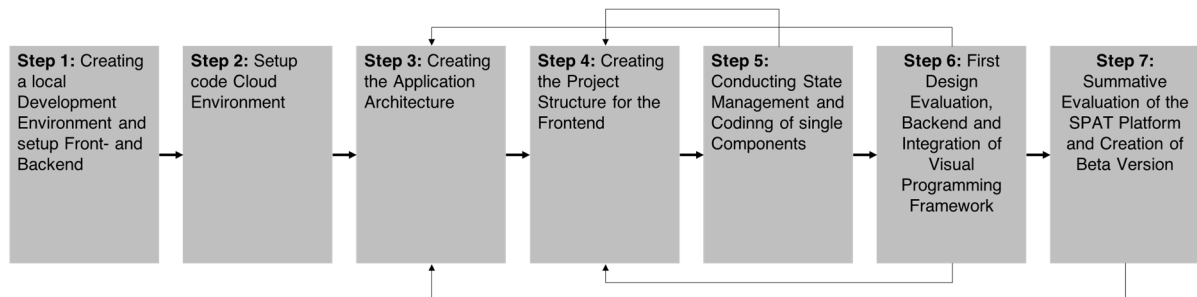


Figure 35. Seven Steps to Create and Test SPAT Platform MyTeachingBot

Source: Own illustration

This section begins with a general overview of the frontend. The introduction page of the web interface (see Figure 36) welcomes the educator with a short explanatory video showing the main benefits of SPAT in education and briefly explains the wizard-like steps of the process model with the help of a two- to three-minute video introduction.

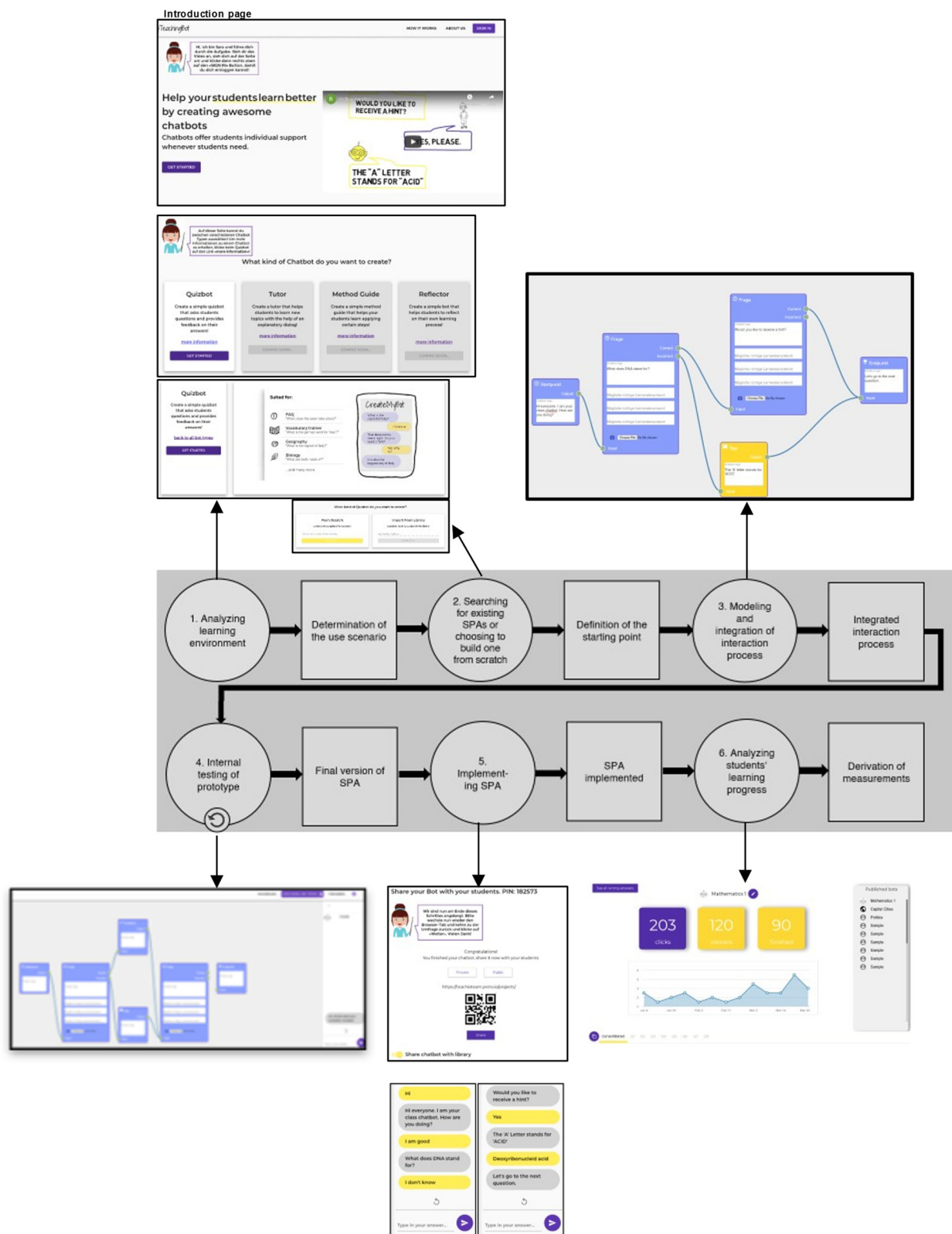


Figure 36. MyTeachingBot SPAT Platform

Source: Own illustration

In step 1 (analysis learning environment), the model provides educators with different use scenarios for creating a SPAT. For example, they can create a SPAT as a Quizbot for computer science topics that challenges students' knowledge

about newly learned concepts. Educators can click on the “more information” button to see example dialogues and to judge whether this type of SPAT is appropriate for their defined learning goals. For more information on the selection of environments, educators can click on “more information”. When educators decide on a use case for their SPAT, they can click on “GET STARTED”. In step 2, educators can decide whether they want to create a SPAT from scratch or search for existing SPAT in the library. Step 3 is the core component of the SPAT platform. When educators begin step 3, they encounter a canvas and can create an interaction flow by using and connecting drag and drop elements. A template with an already implemented interaction flow is displayed in Figure 35. Educators can include different kind of nodes such as question, hint, and feedback nodes. In step 4, after creating the interaction flow, educators can iteratively test their self-created SPAT by typing the questions into the chat window that opens on the right after the SPAT has been saved to the database. They can then start to iteratively make changes to their drag and drop elements until they are satisfied with their SPAT. Finally, educators can share their SPAT via a 5-digit-number, a QR-Code, or a short link with the students. The students can then enter the students’ view mode (a separate application using the same database) and are able to interact with the created SPAT.

As a first step, I set up local and cloud environments. The former is needed to code and test new components or fine-tune the existing code. The latter includes a version control system that helps to keep track of different code versions without creating confusion and conflicts. It also prevents code from being lost by saving it on a cloud. All command line code snippets, which are shown in the following, are based on a Windows OS and the Bash shell.

6.2.3.1 Step 1: Creating a Local Development Environment and Setup Frontend and Backend

To set up the local environment, I used the following software (see Table 39):

Software	Functionality	Recommendation
Code editor	Program to write and organize code on a local machine (e.g., a laptop)	Microsoft Visual Studio Code
Web browser	Application to execute JavaScript (e.g., to display websites)	Google Chrome
JavaScript runtime environment	Program to execute JavaScript on a server	Node

Table 39. Required Software for Programming MyTeachingBot

Source: Own illustration

In the next step, I set up a React application as the frontend. React is a widely used JavaScript library that allows developers to create lightweight user interfaces, also known as single page applications. To initialize this step, I navigated to the desired directory and ran the command `npx create-react-app myteachingbot`.

This command installed all the files required to start coding the React application frontend. Figure 37 shows the initially created files in the Microsoft Visual Studio Code editor.

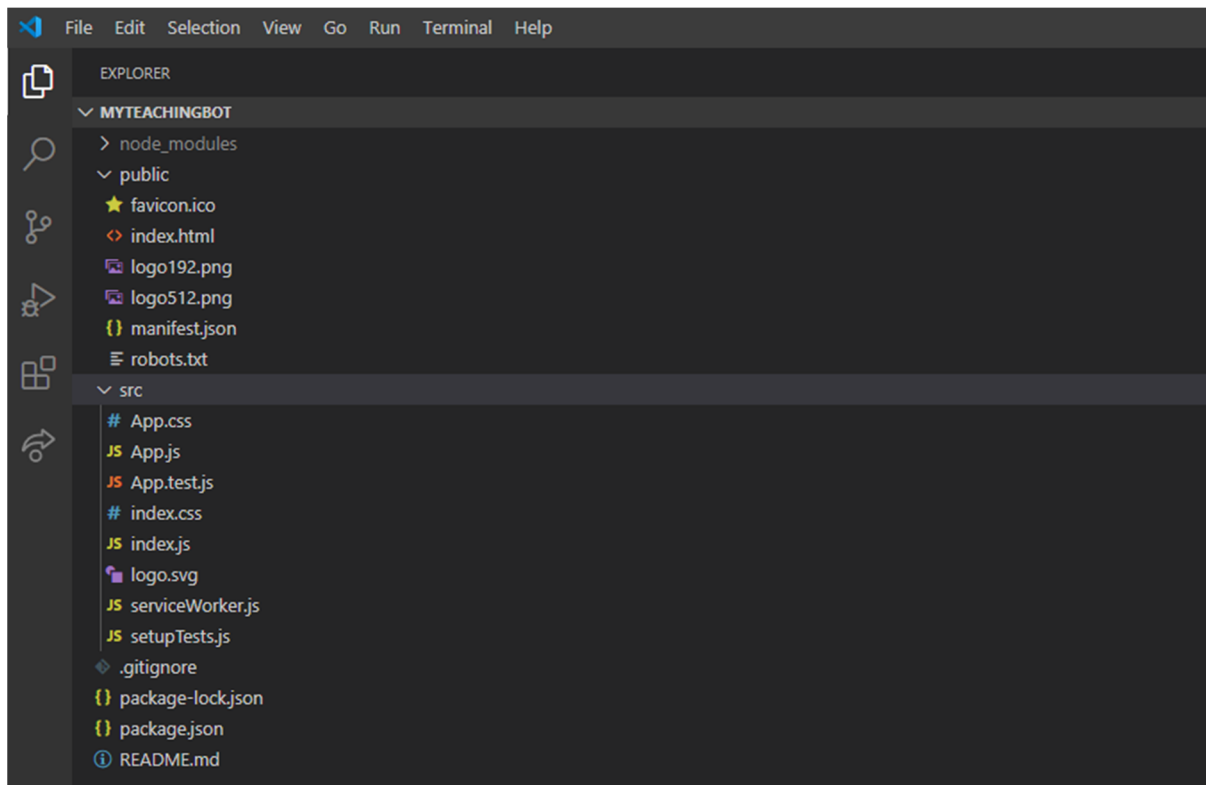


Figure 37. Initially Created File Directory in Visual Studio Editor

Source: Own illustration

In the next step, I directed to the root directory of the application and executed the `npm start` command, which created a server on my local machine and opened a new tab in my Chrome browser from which the application can be accessed by typing <http://localhost:3000>. This helped me to access the web user interface and test the application on my local machine. This last step marked the end of the frontend setup process. To set up the Backend, I downloaded nodeJS and the Express library and also setup the MySQL database server and connected it to the hosting provider (called uberspace).

6.2.3.2 Step 2: Setup Code Cloud Environment

In step 2, I downloaded a version control system and set up the cloud environment. I used the software listed in Table 40 to implement this step.

Software	Functionality	Recommendation
Version control system	Program that tracks every code change made in the past in order to revert to a previous version if needed	Git and GitBash
Code cloud	Program that allows the code to be saved in the cloud	GitHub

Table 40. Version Control Program and Cloud Environment

Source: Own illustration

Git is a version control system that allows reversion to earlier versions if errors occur. GitBash is a command line tool that works in a similar way to the Windows command prompt. In addition to the Windows command prompt, it also includes all Git commands, (e.g., to upload new code changes to the cloud and pull the latest code). For the further process of creating myteachingbot, I used Gitbash throughout. I also created a Git repository on <https://github.com>.

6.2.3.3 Step 3: Creating the Application Architecture

The application consists of multiple Frontend and Backend services that interact together. Table 41 lists the main services and the corresponding layer.

Component	Layer	Status	Description
Educator View (main component)	Frontend	Beta Version built with React	Web interface to empower educators to visually program SPAT and share it with their students. It also makes API calls to the NLP backend.
Student View (secondary component)	Frontend	Beta Version built with React	Web interface that allows students to access the created SPAT via a desktop or mobile device. It also makes API calls to the NLP backend.
SPAT backend	Backend	Beta Version built with NodeJS	Business logic that saves user information, SPAT, and chat information into the database and retrieves it when necessary.
NLP backend	Backend	Beta Version built by Google's Dialogflow	API that handles and classifies incoming user inputs by using advanced machine learning algorithms in order to choose the right responses.

Database	Backend	Beta Version built with MariaDB (a free MySQL compatible Database)	Service that stores all users, SPAT, and chats
----------	---------	--	--

Table 41. SPAT Platform Component Overview

Source: Own illustration

The educator's and students' views are two microservices that are separated to increase security, allow faster scalability, and enable separation of concerns. Figure 37 shows the bigger picture of the architecture and the interaction between the different services.

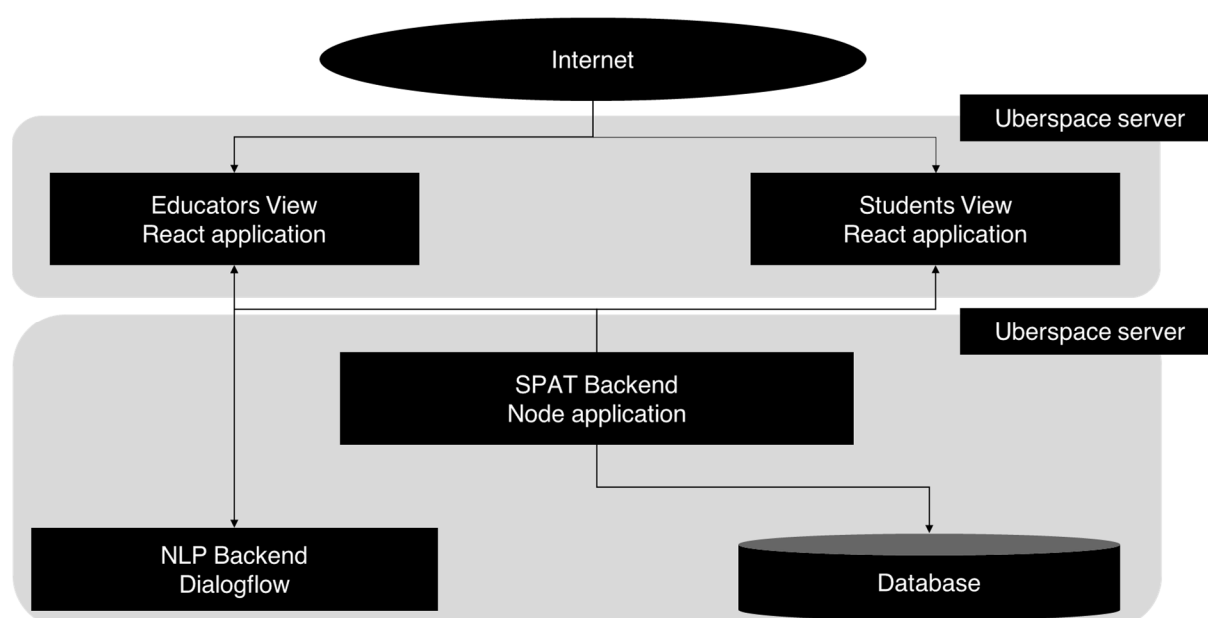


Figure 38. MyTeachingBot Architecture Overview

Source: Own illustration

The frontend and backend are both hosted on an Uberspace server. The educators view (and the students' view) directly interacts with the Google Dialogflow and NLP Backend Dialogflow and also interacts with the Chatbot Backend Node application. The latter writes and reads data in the database.

The educator's view is the main component of "myTeachingBot". Hence, in the following subsections, I focus on the educator's view and only briefly address the students' view. The educator's view includes an app component that contains all other components of the application (see Figure 39). The main goal of the app component is to provide a shell for the different routes on the single page application. Every route typically contains a page component, and each page component typically holds different layout components. Layout components are used to

provide the necessary UI for the educators to interact with the system. In contrast to layout components, state components store internal data that are created during the interaction with the system.

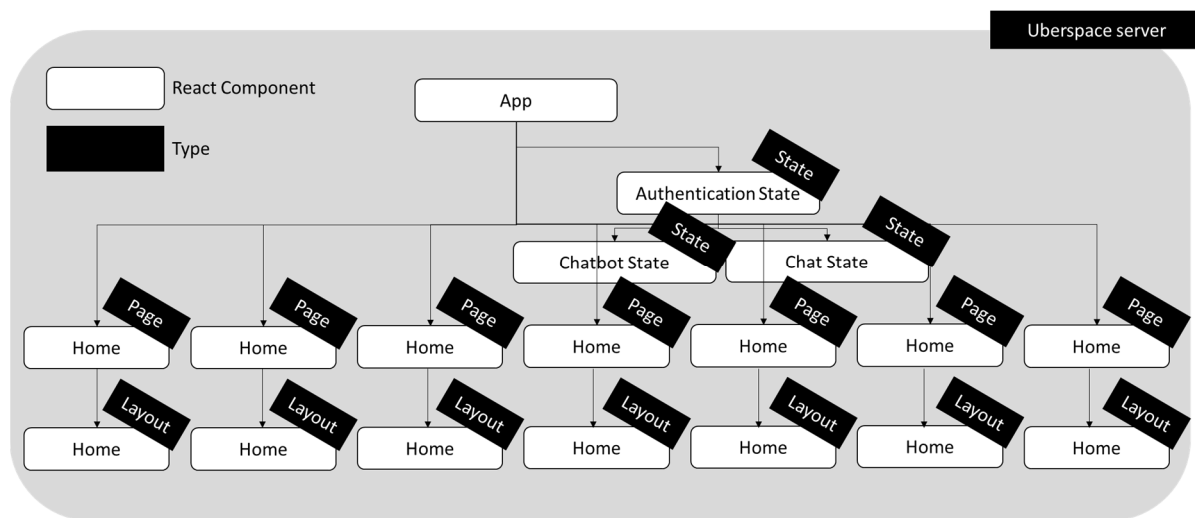


Figure 39. Educator's View Architecture

Source: Own illustration

Figure 39 shows how the different components within the app component work together. This architecture is scalable, allowing new pages to be added if needed. For example, users might want to switch languages to improve their UE. To achieve this is a simple case of implementing a new state component and a new language switcher component, which can be integrated without changing the existing code. As a programming language, I mainly used JavaScript for the web and chose React over Angular and Vue, because React allows the creation of very fast web applications and offers a big community and a high number of online tutorials.

6.2.3.4 Step 4: Creating the Project Structure for the Frontend

The most high-level component of the application is *App.js*. In React, data is passed down and stateful components should therefore be placed in the upper levels of the tree to ensure that child components can access them. I also used Google's Material-UI design library and the corresponding *MuiThemeProvider* component to provide a consistent branding across the entire application. The switch component is required to wrap the page component that provides a shell for the different routes. Figure 40 shows the final components directory.

```

const App = () => {
  return (
    <AuthState>
    <SPATState>
    <ChatState>
    <MuiThemeProvider theme={teechrTheme}>
      <CssBaseline />
      <Router>
        <Navbar />
        <Switch>
          <Route exact path="/" component={Home} />
          <Route exact path="/register" component={Register} />
          <Route exact path="/login" component={Login} />
          <PrivateRoute exact path="/dashboard" component={Dashboard} />
          <PrivateRoute exact path="/createnew" component={Createnew} />
          <PrivateRoute exact path="/choose" component={Choose} />
          <PrivateRoute exact path="/canvas" component={Canvas} />
          <PrivateRoute
            exact
            path="/moreinformationquizbot"
            component={MoreilInformationQuizbot}
          />
        </Switch>
      </Router>
    </MuiThemeProvider>
    </ChatState>
    </SPATState>
    </AuthState>
  );
};

```

Figure 40. Components Directory

Source: Own illustration

Components directory

The React framework logic involves creating different kinds of components and letting them interact with each other. The /Components directory contains subdirectories for /Pages and /Layouts and /Rete. Rete is a visual programming framework that I used to visualize the drag and drop logic of the interaction flow.

Context directory

The context directory allows information to be saved on the frontend and made accessible to other components in the application without the need to interact with the backend. The concept of contexts allows developers to make their React application “stateful” and introduce the concept of a “global state”. We created a global state for the authentication process, the SPAT meta data, and the Chat meta data. The authentication context allows any child component to access objects such as the authentication token or functions such as the login, logout, or fetching user data. The SPAT context allows child components to access the metadata of a created SPAT. For example, the dashboard child component needs to access certain user SPAT metadata and user information to render a list of already created SPAT belonging to the logged-in user.

6.2.3.5 Step 5: Conducting State Management and Coding of Single Components

A state stores data on the frontend—the web browser—without needing to communicate with the backend. This makes an application much faster. The local state describes the state within a component, and the global state describes the state at global level—for example, the authentication state (see Figure 39). An example of a component-level state is the current status of a modal component. In contrast, an application-level state can be the authentication status (logged in or logged out) that needs to be accessible for multiple page and layout child components. Below, I present an example application of the local state.

Local state

One example for a component-level state is the Choose.js component (see Figure 41) where users give their SPAT a name.

Figure 41. Component with Local State

Source: Own illustration

The SPAT name is not required in any other component, and it is therefore sufficient to deal with it in the local state. Implementing the local state involves three steps.

Initiate local state within the Choose.js component with the help of the useState hook (see Figure 42).

```
// local state: create state for input field

const [newQuizbotState, setnewQuizbotState] = useState({
  quizBotName: "",
  helperTextState: "",
  errorState: false,
});

const { quizBotName, helperTextState, errorState } = newQuizbotState;
```

Figure 42. Initializing of Local State

Source: Own illustration

The useState hook returns an object storing the state and a function that can be called to change the state object.

Linking the quizBotName input field value to the quizBotName property of the state (see Figure 43).

```

<TextField
  className={textField}
  id='createNewQuizbot'
  name='quizBotName'
  type='text'
  label='Gib deinem Quizbot einen Namen'
  placeholder='Gib deinem Quizbot einen Namen...'
  helperText={helperTextState}
  error={errorState}
  value={quizBotName}
  onChange={quizBotNameHandler}
  fullWidth
  />

```

Figure 43. Link of TextField Input Value to the QuizBot Name Property

Source: Own illustration

The value property of the input field component is linked to the local state. Whenever the component is manipulated (e.g., when the user types something), the onChange property of the component calls a function that should change the value of the state and the input field.

Updating the quizBotNameHandler function to change the state when the input field updates (see Figure 43).

```

// function: add whatever is typed in to local state

const quizBotNameHandler = (e) => {
  setnewQuizbotState({
    ...newQuizbotState,
    [e.target.name]: e.target.value,
  });
};

```

Figure 44. Updating the QuizBot Name Handler

Source: Own illustration

The function shown in Figure 44 induces that the value within the input field is immediately transferred to the property in the local state object.

6.2.3.6 Step 6: First Design Evaluation, Backend and Integration of Visual Programming Framework

I conducted an initial design evaluation, in which I showed the alpha version of the SPAT platform to three educators. The findings of this evaluation made it necessary to add the following main functionalities and features into MyTeachingBot, essential components in empowering non tech-savvy educators to create their own SPAT:

1. Server backend with Axios library
2. Consistent theming (using the Material-UI library)
3. Visual programming framework (using the Rete library)

The goal of this subsection is to explain the core concepts and workings behind these features.

Server Backend with Axios Library

The use of local and global states allows information to be stored on the browser. This way of storing data is temporary, and the data will be lost as soon as the web user interface is reloaded. This will not be sufficient when bringing the SPAT platform to production—a new user and SPAT should be saved permanently so that the user can access the SPAT from different devices. Figure 45 shows how my application communicates with the backend. For a better overview, I defined all functions with an API Call in the State (e.g., Chatstate.js) components.

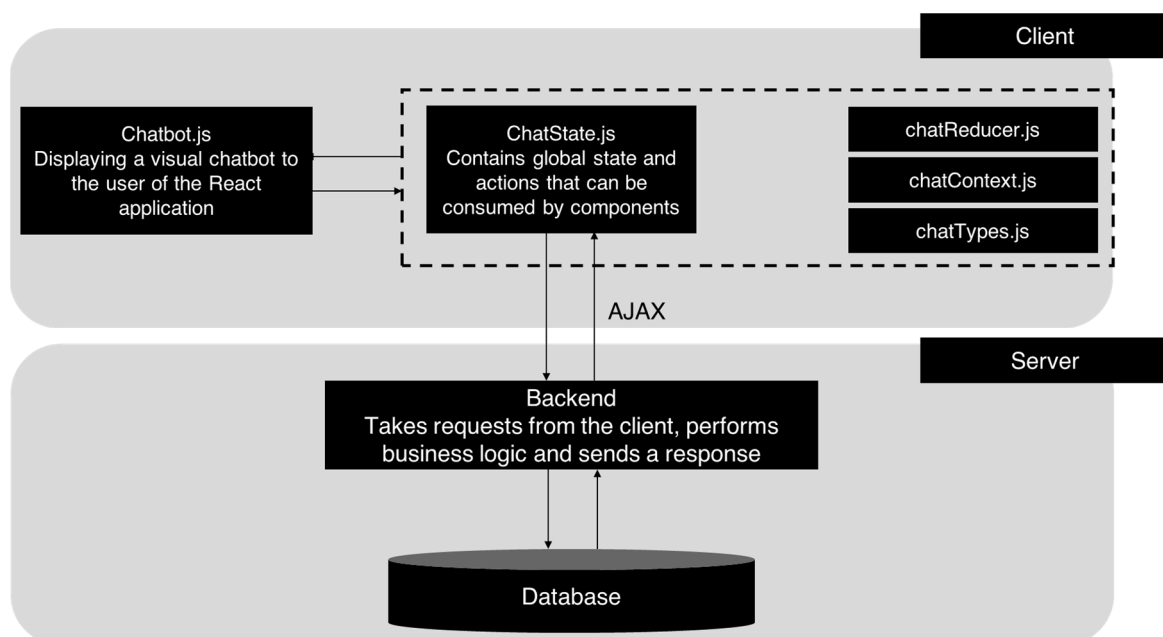


Figure 45. Client-Server Communication in Educator's View

Source: Own illustration

I used the AJAX library to perform HTTP requests.

In the next step, I created the backend of the application. For the backend, I decided to split the project folders into config, controllers, models, and routes directories. In the config directory, I defined the general backend path and saved the database credentials to connect the database. Moreover, I decided to use a Model-Controller design pattern that divides the related program logic into two interconnected elements (the controller and model directory) in order to separate internal representations of information from the ways information is presented. In this manner, I separated the SQL queries in the models directory from the function declarations in the controller directory. Consequently, if I want to change the database in the future, I only have to change code within the models section. The routes directory defines the different resources that can be accessed via the frontend (e.g., to create a new SPAT). Appendix C shows examples of code snippets of the route, controller, and model directory for creating a new SPAT.

Consistent Theming (Using the Material-UI Library)

To provide a unique branding and a uniform design, I used the Material_UI library from Google. I imported the desired Material-UI components, created a customized MUItheme, and added extra CSS to the component using the React CSS syntax. I used the Hook API, where I first created the style with the “makeStyles” command and imported it in the functional component.

Visual Programming Framework (Using the Rete Library)

Rete.js is a JavaScript framework for visual programming, which I used to create the drag and drop editor necessary for creating the interaction dialog. Figure 46 shows how this framework was used and visualized in my editor.

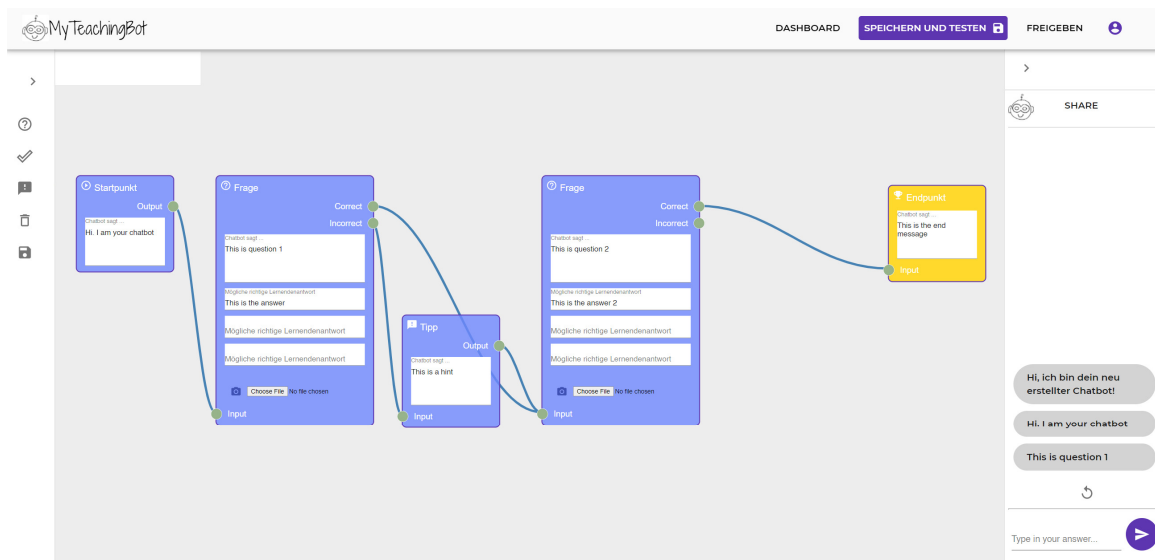


Figure 46. SPAT Editor

Source: Own illustration

Figure 47 shows the different parts of the ReteJS framework and how they were integrated into the application.

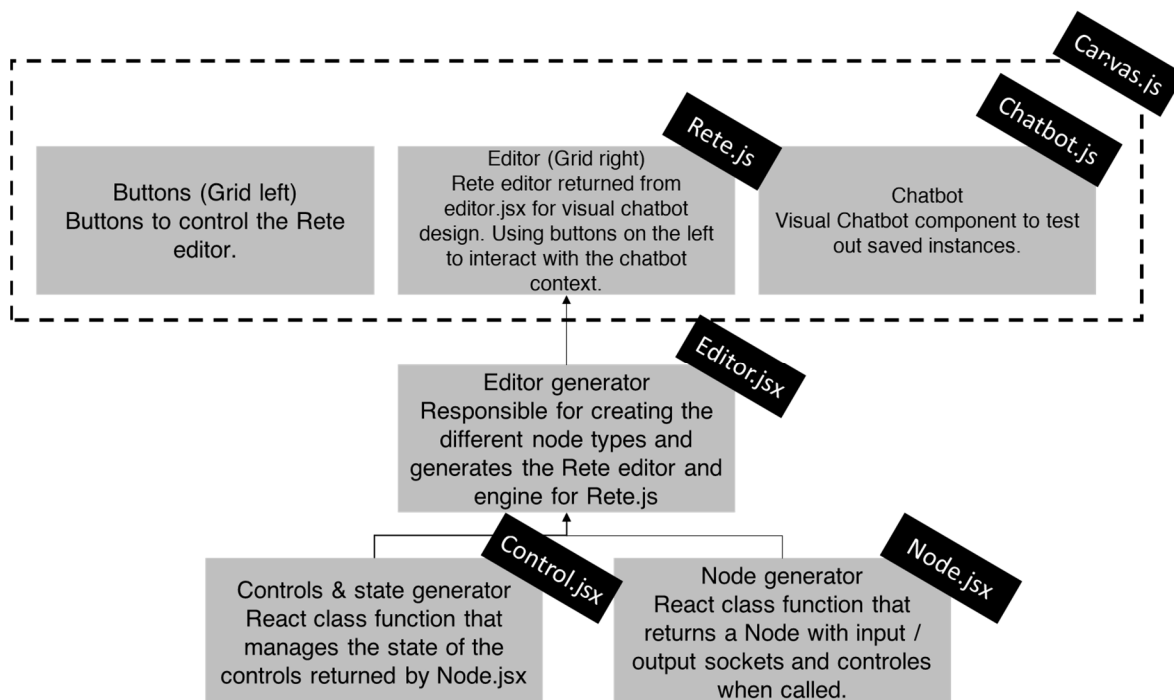


Figure 47. ReteJS Integration into Educator's View

Source: Own illustration

According to the identified theoretical and practical requirements, educators need different kinds of nodes to create their interaction flow. Specifically, I identified a start node, question node, feedback node, hint node and an end node block to

create a useful dialog. Educators can combine these node blocks in any way that makes sense to them, providing they start with the start node and end with the end node. The question node block consists of the SPAT's question and possible correct answers from the student. The feedback node block consists of a single input field that provides the SPAT's answer when a student has answered the question correctly. The hint node block offers assistance whenever the students does not know the answer.

6.2.4 Summative Evaluation for the SPAT Platform (Step 7)

6.2.4.1 Sample

My application domain involves educators across different levels. The main goal is that non-tech-savvy educators become empowered to create and embed SPAT in their own blended and online learning environment without the need for specific technical skills such as programming experience. The evaluation was conducted by 38 educators. Moreover, I conducted 12 UE interviews with the educators after they created SPAT with my SPAT Platform MyTeachingBot. Table 42 shows the data relating to the educators who participated in the evaluation.

Interview-ees	Characteristics
Future Educators (N=25)	Average Age: 24.12; level of education: 8 x primary school, 6 x secondary school, 9 x high school, 0 x vocational school, 2 x university Nationality: 5 x Swiss, 16 x German, 1 x Italian, 1x Icelandic, 1x Austrian, 1x Chinese, gender: 16 women, 9 men
	Type of school and subject: 7 x mathematics, 4 x chemistry, 7 x German, 1 x history, 10 x English, 3 x everyday culture and health, 3 x politics and economy, 2 x sports, 2 x ethics and religion, 2 x social sciences, 2 x geometry, 1 x French, 1 x biology, 1 x physics, 1 x banking and finance, 1 x business innovation, 1 x business studies
Educators (N=13)	Average Age: 31.62; level of education: 5 x primary school, 1 x secondary school, 3 x high school, 1 x vocational school, 3 x university Teaching experience: 0 to 30 years; nationality: 10 x Swiss, 2 x German, 1x Italian; gender: 6 women, 7 men
	Subjects: 7 x mathematics, 1 x biology, 1 x physics, 1 x chemistry, 3 x English, 6 x sports, 5 x music, 3 x ethics and religion, 1 x history, 1 x geography, 1 x Italian, 2 x media and informatics,

	2 x nature-human-society, 4 x German, 4 x design, 1 x business and law, 1 x business studies, 1 x architecture, 1 x history of art, 1 x all subjects of primary school
--	--

Table 42. Sample Characteristics

Source: Own illustration

6.2.4.2 Process

The second formative evaluation was divided into two and three parts, respectively.

In the first part of the evaluation, the probands recreated a SPAT with the help of a given conversational dialog. The predefined dialog is depicted in Figure 48.

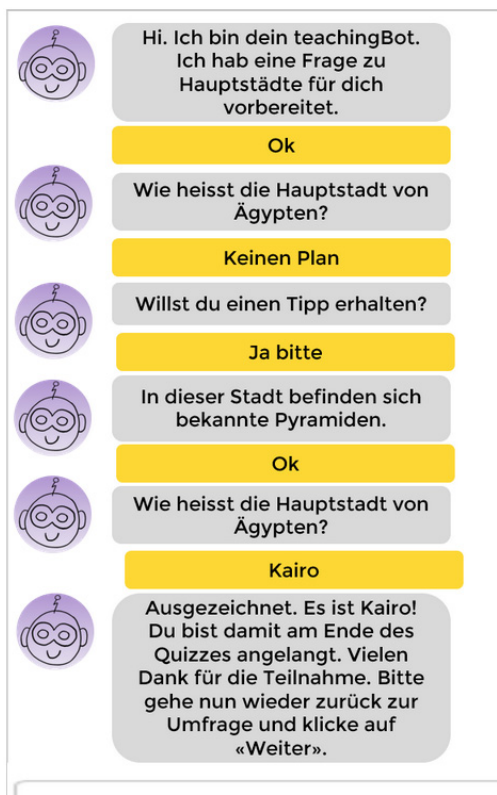


Figure 48. Predefined Dialog for Conversational Dialog

Source: Own illustration

After the first part of the evaluation, the educators created their own SPAT according to their own ideas and without a predefined conversation script. With the help of a screen recording tool, I analyzed both result- and process-oriented outcomes where improvement potential exists. I analyzed educators' behavior on the platform with a screen recording tool called rrweb. For the second part of the evaluation, participants had to complete a survey. In addition to demographic data, questions primarily concerned the content of the application of the SPAT. This included questions in the areas of Personal Innovativeness, Interactional

Enjoyment, User Satisfaction, Intention to use, Perceived Usefulness, Ease of use, and UE Test (the exact constructs can be found in Appendix C.4). Furthermore, I created open questions to gain more detailed answers and USs to optimize the SPAT. I posed these questions: How do you rate the tool? Could you imagine integrating the SPAT in your future teaching sessions? Which requirements and expectations would have to be given in order for the SPAT to be used by the students? Do you have general suggestions for improvement of the platform? What else would you wish for? If you had difficulties in handling the SPAT, what exactly did you find difficult and what would you improve?

The interview questions aimed at ascertaining the concrete UE of the educators when using the platform. In the interviews, I performed the individual steps of MyTeachingBot again and asked the educators about their impression of the step. For example, I asked: What did you like most about this specific step? How would you start building the SPAT? Is there any information missing? How easy is it to develop the SPAT dialogue?

6.2.4.3 Quantitative Results

Table 43 shows the descriptive and analytical results of my measured constructs. On average, it shows that all results are under the mean (where 1 is best and 7 is worst). The one-sample t-test reveals that interactional enjoyment, user satisfaction, ease of use, and the UE test all gained positive results.

Constructs	Evaluations (1 is best, 7 is worst, 4 is the median)	T-Test
Interactional enjoyment	3.445	0.0145
User satisfaction	2.66	0.0001
Intention to use	3.59	0.0581
Perceived usefulness	3.86	0.5123
Ease of use	2.57	0.0001
User Experience Test	2.74	0.0001

Table 43. Quantitative Results of Second Formative Evaluation

Source: Own illustration

6.2.4.4 Qualitative Results (Change Clusters)

I developed the following CRs from the educators' USs based on the screen recordings, the qualitative answers in the survey, and the UE interviews.

User Stories: Educators	CRs	Source
USE1: As an educator, I want to intuitively understand what I have to add into the canvas fields	CR1: The functioning of the canvas structure should be more intuitive.	Videointerview
USE2: As an educator, I want to be able to use more space to create a Canvas	CR2: The canvas should take up more space on the screen in order for the canvas to be designed in an effective way. Moreover, the zoom function should allow the switch to fullscreen.	Videointerview, screen recording
USE3: As an educator, I want to be able to use SPAT from other educators and individualize them	CR3: MyTeachingBot should allow educators to publish their SPAT for other educators.	Videointerview
USE4: As an <i>educator</i> , I also want the SPAT to recognize similar answers as right answers	CR4: MyTeachingBot should integrate a more sophisticated NLP API instead of simple word-pattern matching algorithms.	Videointerview, Online survey
USE5: As an <i>educator</i> , I want to create more complex conversations.		Videointerview
USE6: As an educator, I want to switch the language to German.	CR5: MyTeachingBot should be available in different languages.	Videointerview
USE7: As an <i>educator</i> , I want to be able to see my students' statistics	CR6: MyTeachingBot should include a function where educators can track their bots and see how the students did.	Online survey, Videointerview

USE8: <i>As an educator, I want to create a SPAT in a short amount of time.</i>	CR7: MyTeachingBot should be straightforward without too many additional functions.	Videointerview, online survey
--	--	-------------------------------

Table 44. Qualitative Results and Corresponding Change Clusters

6.2.5 Refinement of Platform and Third Evaluation

Based on these CRs, I created a beta version of MyTeachingBot that addresses the CRs. The biggest changes and corresponding CRs are briefly listed below, and they were also included in the initially presented overview of the application in Subsection 6.2.3:

- Inclusion of Google's Dialogflow API (CR4)
- Redesign of canvas interface (CR1 and 2)
- Including a SPAT database (CR3)
- Including a language switcher (CR5)
- Inclusion of a statistics dashboard (CR6)
- Pages that lead to summarized canvas (CR7)

6.2.6 Contribution

In this study, I developed a SPAT platform based on the process model described in Section 6.1 that empowers educators to create and embed SPAT for their own online and blended learning environments. This SPAT can adopt the role of a private tutor by responding adaptively to the personal needs of students from diverse backgrounds. The study shows that educators are satisfied with the platform and that they intend to use this platform for creating SPAT for their own learning environments. I have thereby addressed the last RC3 and RQ3 and tested DP5, accessibility. Accessibility is a necessary next step in SPAT research, as it is essential for helping the technology enter learning environments and for leveraging its full potential. In the next section, I summarize the overall contributions of this thesis and highlight exciting research questions linked to the theme of this dissertation.

7 Summary of Contributions and Areas for Future Research

“Nothing is as practical as a good theory (Lewin 1943, p. 129).”

The discussion of the studies detailed in the previous chapters has already given micro-perspective insights into the contributions of this dissertation. The goal of this section is to widen the perspective a little and show how the individual studies contribute to addressing the challenges I outlined at the beginning of this dissertation. Hence, this section provides a coherent overview of the individual studies' contributions to theory and practice and on this basis provides an outlook for fruitful avenues for future research areas not covered in this dissertation.

7.1 Theoretical Contributions of the Dissertation

I discuss the theoretical contributions of my study in the context of certain criteria related to the philosophical foundations of the dissertation. Weber et al. (2012) created a framework and corresponding dimensions that can be used to evaluate the quality of theories within the information system discipline: Importance, Novelty, Parsimony, and Level.

The first criteria, the *importance* or *utility* of a theory, is often assessed via judgments made about the importance of its focal phenomena (Corley and Gioia 2011) and whether it is considered as important from the viewpoint of research and from the viewpoint of practice. The main focus of this study is how to offer individual support to students in online and blended learning environments. Many studies from the last two decades have highlighted the importance of this issue (Xie et al. 2019). This dissertation divides the matter into three stated RCs that constitute the problem statement of this dissertation (see section 1.1). To find solutions for these RCs, I formulated research questions that focus specifically on SPAT to address the overall issue of offering individual support in online and blended learning settings.

The second criteria, *novelty*, relates to the extent to which a theory is novel. A theory will be deemed novel to the extent that it changes the paradigms used by researchers to investigate the phenomena within their discipline (Devlin and Bokulich 2015). It provides a way of resolving problems within their discipline—that is, empirical observations of phenomena that existing theories are unable to explain or predict (Weber 2012). In the context of design science research, Hevner and Gregor (2013) defined improvement, exaptation, and inventions as novel knowledge contributions. This dissertation describes a design theory that can be categorized as improvement, and does so by using a new class of ITS (SPAT) to address the problem of offering effective individual support in online and blended learning environments. Until now, other classes of IT systems have not been able to improve learning processes and increase learning success to the same extent.

The third criteria, *parsimony*, states that theories are parsimonious, which means that they achieve good levels of predictive and explanatory power in relation to their focal phenomena using a small number of constructs and associations. By using a small number of constructs, they limit the size of their conceivable state space and conceivable event space. Thus, the boundary of parsimonious theories can be stated more precisely, because the class of things, associations, states, and events covered by the theory can be defined precisely. Based on the TML framework proposed by Gupta and Bostrom (2009), I used a very small number of constructs to analyze the effect of SPAT on students' learning processes and on the learning outcomes. I investigated this effect on a variety of different levels (high school, vocational business school, higher education) and contexts (online and blended learning environments) and used different experiment settings (controlled laboratory settings and real-life field experiments) to strengthen the explanatory and predictive power related to the effectiveness of SPAT.

The fourth criteria, *level*, relates to the level of the theory. Some theories cover a very narrow, constrained set of phenomena ("micro-level" theories), while others cover a broad range of phenomena ("macro-level" theories). Merton (1968) argued that the primary theories used in a discipline ought to be "middle-range" (or "meso-level") theories. On the one hand, such theories avoid "narrow empiricism." On the other hand, they avoid being so general that it is difficult to test them empirically. Moreover, meso-level theories often have value because they link the micro-level world and macro-level world in a discipline. While the individual studies' constructs and associations might be very narrow in terms of micro-level theories, the combined studies might have a high level of generality and can therefore be called a "meso-level" theory. My study provides broad, overall insights into how a new class of ITS (SPAT) must be designed to address the problem of individual support in online and blended learning environments.

The studies in this dissertation are influenced by three major philosophical worldviews (see Section 3.1). My studies concerning the experiments (see Chapter 5) are mainly situated in the quantitative paradigm and mostly influenced by the perspective of critical realism. Thus, the truth of the contributions has to be considered carefully, because they are only empirically tested and it is not possible to test every single circumstance (see Popper's falsifiability principle (Popper 1959)). Complementing the quantitative research paradigm, some of the studies also use the interpretivist perspective of the qualitative research paradigm. These studies sought to construct knowledge as accurately as possible through the continuous and rigorous interpretation of interview data (see section 4.2) and focus group data (see section 5.2). Hence, standards concerning falsifiability other than those in the quantitative studies have to be considered. Finally, in regard to the overarching design science research approach of the dissertation, the work is influenced by both the pragmatic perspective (e.g., what brings utility is good)

and the quantitative paradigm with its critical realist view. The latter view is important in this study, since theoretical propositions and relevant hypotheses concerning SPAT design decisions are tested with empirical data throughout the dissertation. Given that the design theory in this thesis is based on a limited number of design science research studies, the derived DPs and the corresponding design theory cannot be considered as “the truth” and should be treated with caution. The following subsections outline the theoretical contributions derived from addressing the research questions, which can be summarized with the help of Gregor and Jones’ (2007) anatomy of a design theory framework.

7.1.1 Aspects of Smart Personal Assistant Technology that Affect Learning Processes and Learning Outcomes

The first overarching RQ in my dissertation was:

RQ1: What are relevant DPs of Smart Personal Assistant Technology in online and blended learning environments?

The first RQ was divided into two sub-RQs that were addressed by a literature review and then by an interview study. I aimed to answer RQ1a in section 4, this question being as follows:

RQ1a: What is the state-of-the-art regarding Smart Personal Assistant Technology in education?

The results of the literature review detailed in Section 4.1 show the current state of the art of SPAT from a structure (input), learning process (process), and learning outcome (output) perspective. The review included 80 articles that helped me to propose the following four contributions.

First, based on a TML input perspective, the findings of the review revealed that students’ individual differences are decisive in how they perceive interactions with SPAT. However, no study was able to prove that individual differences have a significant impact on learning outcomes. Furthermore, there are still many unanswered questions about how SPAT can be designed to address these individual differences. In the pretest described in Section 5, I tried to address this research gap by investigating the influence on learning outcomes of several student characteristics such as personal innovativeness, learning style, gender, nationality, and age. None of the mentioned constructs had a significant influence on learning outcomes in this study. Consequently, the study cannot confirm these review-based results.

Second, based once again on a TML input perspective, the findings show that SPAT can be characterized by its construction method (retrieval based or generative), its interaction mode (text and/or voice), and its intelligence (context-adaptive or non-adaptive). The review shows that for closed domains with a fixed set of possible questions, simple rule-based SPAT might be more than sufficient. In all other cases, SPAT with artificial intelligence techniques lead to better learning

outcome results. The design and use of this latter type of SPAT are still under-researched. Hence, in all my studies I focused on building SPAT that uses machine learning techniques and also investigated the design and use of voice-based SPAT, as detailed in Sections 5.2 and 5.3.

Third, based on a TML process perspective, the review revealed that interaction with SPAT is fundamentally different to interaction with educators. Studies showed that students apply simpler sentences, lack a rich vocabulary, and get distracted more easily. This dissertation confirms this research stream by identifying interaction patterns that are different to those associated with human–human communication (see Section 5.1). This revives the more general discussion about the role of SPAT and other AI systems in education, and the question of whether AI systems should be an *educator-replacing* or an *educator-assisting* technology. The role of AI used only for automation in education systems is being questioned (Baker 2016). Cukurova (2019) argues that AI systems should be considered as part of a continuum to the extent that they are decoupled from humans, rather than a means of providing full automation. Based on the results of my experiments, I argue that SPAT can be positioned on both sides of the continuum. In my studies (e.g., see Section 5.3), SPAT has served mainly as an educator replacement to aid educators in settings where they cannot provide individual support due to financial and organizational reasons. SPAT could therefore be used mainly in isolation from educators. In addition, SPAT could also be used to support the decision-making processes of educators. When students interact with SPAT, a large amount of student data can be collected, such as the current level of knowledge, frequency of exercises, and motivation at individual and collective levels. This student data can help educators to introduce new measures (e.g., to explain content again, to offer low-achievers individual help). I believe the question about the role of AI systems in education is less about an either/or solution and more about both—that is, replacing educators where they cannot offer help and informing educators’ decision processes. An important research gap to be addressed in the future is thus how and where human or SPAT assistance is preferable and how educators can use SPAT. This dissertation compared SPAT assistance with human assistance, as described in Section 5.1, and created a SPAT platform that should help educators in creating their own SPAT for their online and blended learning environments.

Fourth, based on a TML output perspective, the review can be used to improve different kinds of learning outcomes such as affective, cognitive, and metacognitive learning outcomes. For example, studies showed that SPAT can help students to gain factual knowledge (Ruan et al. 2019). Moreover, SPAT can actively control and regulate learning processes (Duffy and Azevedo 2015). Another study showed that SPAT can be used to increase student motivation in learning and strengthen their self-perception (van der Meij et al. 2015). There is still very little empirical evidence about how these systems can foster different kinds of learning outcomes. This dissertation addressed this issue by providing empirical evidence

for different kinds of learning outcomes. In the study described in Section 5.2 I focused on problem-solving skills as part of the metacognitive skillset, and in the study described in Section 5.4 I focused on cognitive learning outcomes.

To summarize, in addressing RQ1a it became apparent that there is still limited knowledge regarding how SPAT can be designed and how it may influence the learning process and various learning outcomes. This study's literature review is strongly connected with RC1 (limited knowledge about the use and requirements of SPAT) and forms the basis of my second sub-question, which was a prompt to gathering requirements and proposing a first set of DPs for SPAT with the help of an interview study:

RQ1b: What are relevant design principles of Smart Personal Assistant Technology?

I addressed this RQ in Section 4.2, which describes the gathering of requirements and the derivation of the DPs. The literature revealed that there is still insufficient knowledge of how educators and students use SPAT. This is why I asked five educators and 20 students if and how they would use SPAT in their courses. Furthermore, I scanned three different theoretical perspectives (educational, technical, and voice user interface perspective) to gather requirements. Based on the "Anatomy of a DP" framework proposed by Gregor et al. (2020), I proposed a first set of relevant DPs on how to design SPAT in online and blended learning environments. The most relevant DPs I identified were humanness, conversational instruction, instructional scaffolding, an engaging dialog flow, and accessibility. This first set of DPs was further tested by addressing RQ2, which involved a series of laboratory and field studies.

7.1.2 Design of Smart Personal Assistant Technology in Online and Blended Learning Environments

The second research question of this dissertation is:

RQ2: How can Smart Personal Assistant Technology be designed to improve learning processes and learning outcomes in online and blended learning environments?

I addressed this question by answering two sub-questions that focused on different DPs and different learning environments. The first sub-question was about blended learning environments, and the second research question was about online learning environments. The first sub-question was as follows:

RQ2a: How can humanness and conversational instruction incorporated into Smart Personal Assistant Technology improve students' learning success in blended learning environments?

I conducted two different kinds of experiment to answer this sub-question. Each of the experiments tested one of the DPs (humanness and conversational instruction) in a blended learning environment. Thus, each experiment provides a major contribution to the discussion concerning the design and use of SPAT in online and blended learning environments.

The first study concentrated on the humanness DP, where I was able to show that SPAT that includes human-like characteristics is even able to outperform scripted human tutors and involves unique interactions. Based on this finding, in a second study I incorporated conversational instruction (DP2) and tested the usefulness of SPAT in a real-life setting. I created a SPAT for a second-year high school and vocational business school class to conduct four homework assignments and compared the learning outcomes with those of the control class. The data showed that students used the SPAT on a regular basis during the five-week experiment and revealed that SPAT is able to improve skill development. Significantly, the findings also indicate changes in students' learning processes. Students reported that it felt like having their own coach and that they were also more engaged when learning in environments that are normally not used as learning spaces (e.g., the living room).

RQ2b: How can dynamic scaffolding and an engaging dialog flow incorporated in Smart Personal Assistant Technology improve students' learning success in online learning environments?

After testing the first two DPs, I continued via RQ2b by testing DP3 (dynamically scaffolding) and DP4 (engaging dialog flow) in an online learning environment. To test DP3 in an online learning environment, I first conducted a laboratory experiment. The laboratory experiment helped me to control otherwise disturbing variables (no randomization, etc.) and make additional measurements. This time, I investigated whether dynamically scaffolding mechanisms have an effect on student engagement and cognitive learning outcome (DP3). Based on galvanic skin response detectors and facial expression scores, I was able to prove that there is a significant correlation between dynamically scaffolding SPAT and learning outcomes in an online learning environment. In the second study linked to RQ2, I focused on DP4 and implemented an engaging dialog flow in a SPAT that I used within an online learning course. The SPAT assisted students in learning how to conduct a SWOT analysis and compared the results with a traditional textual learning aid. In a post-test task, this study also showed significant results for the experiment group that used SPAT compared to the control group in terms of learning outcomes. The significance shown in this study strengthens the knowledge gained so far and shows that the four DPs of RQ2 are likely to enable SPAT to positively influence the learning process in order to achieve learning success.

To summarize, the insights derived from answering RQ2 relate to a theory of explanation and prediction as well as design and action (Gregor 2006). First, the

empirical evidence from the experiments explains and predicts how well- designed SPAT can influence students' learning processes and different kinds of learning outcomes. Second, the results show how SPAT can be designed to impact learning processes and learning outcomes in online and blended learning sessions. Both theories contribute to the body of knowledge of SPAT in TML research.

The final DP to test was accessibility. After demonstrating the usefulness of SPAT in online and blended learning environments in Section 5, it is necessary to make this technology accessible for educators so that they are empowered to adopt and use SPAT in their own classes. The overarching RQ3 tackles this challenge.

7.1.3 Empowering Educators to Create their own Smart Personal Assistant Technology

The third RQ of my dissertation is as follows:

RQ3: How can non-tech-savvy educators design Smart Personal Assistants for their own online and blended learning environments?

This RQ3 was divided into two sub-RQs, grounded in a theory- and practice-driven design science research approach. The first sub-RQ was a prompt to derive requirements from theory and practice to create a first process model of how educators can create SPAT for their own online and blended learning environments. The second sub-RQ addressed instantiating the process model into a usable web user interface as a SPAT platform.

RQ3a: What are the relevant steps that help non-tech-savvy educators to create SPAT for their own learning environments?

To answer this sub-RQ, I gathered requirements from educator and student interviews. I then gathered requirements from theory using different perspectives. I was then able to propose a first process model and evaluate it by accompanying educators while creating SPAT. The process model provides several contributions relating to a theory of design and action. Current design knowledge in the field mainly addresses educational software designers (Graesser et al. 2018a). My process model is aimed at non-tech-savvy educators. My findings show how the design of SPAT can be embedded into a socio-technical context such as learning environments and can provide initial assistance to educators seeking to create their own SPAT and embed it in their own online and blended learning environment. The result is a first attempt to bridge the gap between the technical knowledge of software designers and the learning knowledge of educators.

The process model alone cannot yet enable the educator to design SPAT and make it available to the students. This is why I created a web user interface as a SPAT platform, which also helps to instantiate and evaluate the process model. The second sub-RQ is thus as follows:

RQ3b: How can non-tech-savvy educators design Smart Personal Assistants for their own online and blended learning environments?

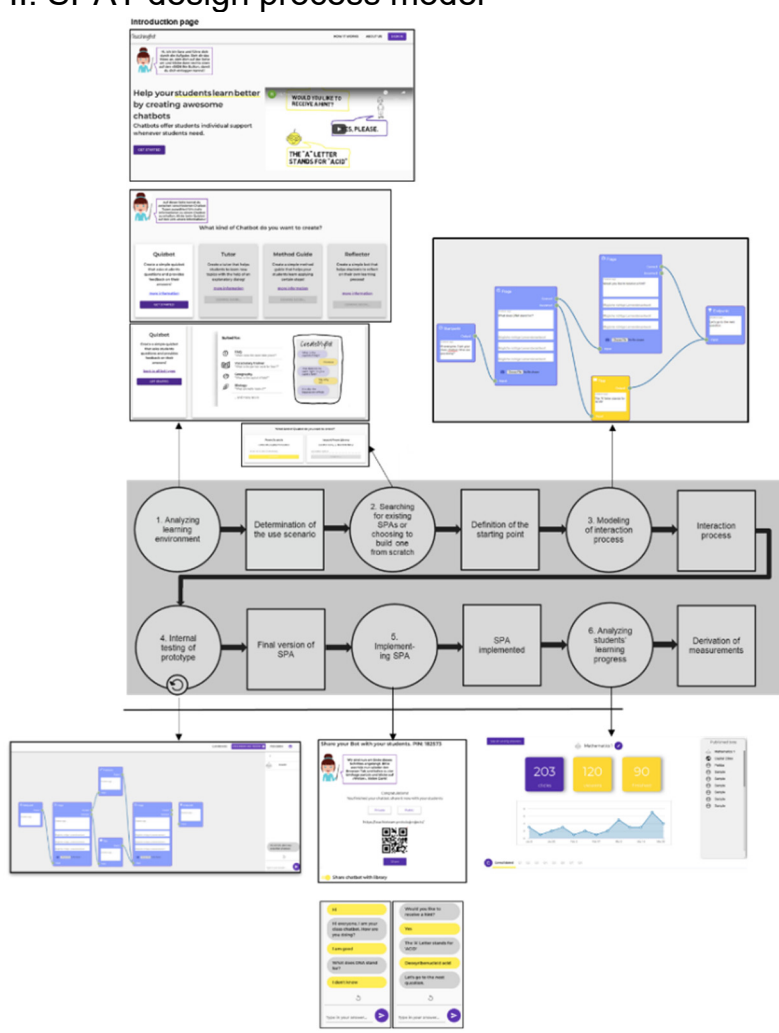
To answer RQ3b, I created a SPAT platform where non tech-savvy educators can create SPAT for their own online and blended learning environments. The design of this platform is strongly connected to the process model developed by addressing RQ3a. It helps educators to conduct the steps of the process model. The web user interface provides several contributions relating to a theory of design and action. First, the web user interface extends the research stream of authoring tools for tutoring systems by showing what such a tool would look like for a new subtype of SPAT. Second, it closes the gap between the technical knowledge of software designers and the learning knowledge of educators. Educators are thus empowered to create the technology for their own use and are therefore more confident in using it.

To summarize, by addressing RQ2 I was able to show that carefully designed SPAT can have a positive effect on the learning process and learning outcomes. The work associated with RQ3 built on this insight and helped me identify details that make the design and development of SPAT accessible to non-tech-savvy educators. As a result of these developments, educators will be more likely to adopt SPAT and use it in their learning environments and will be able address the issue of individual support in these settings. RQ1, RQ2, and RQ3 together form my SPAT design theory, which I will present with the help of the “anatomy of a design theory” framework proposed by Gregor and Jones (2007).

7.1.4 Components of a Design Theory for Providing Individual Support in Online and Blended Learning Environments with the Help of Smart Personal Assistant Technology

Gregor and Jones (2007) introduced a framework consisting of eight fundamental components that I explained more precisely in Subsection 3.6.2. Table 45 summarizes the findings of the individual studies and combines them to form a design theory.

	Type	Component examples
(1)	Purpose and scope	The aim is to formulate DPs for SPAT resulting in a platform that enables educators to create properly designed SPAT by themselves in order to address the problem of individual support in online and blended learning environments. The DPs and the platform are developed for non-tech-savvy educators at different educational levels and in different subjects. The platform is inspired by the DP1-4 and considers a number of parameters that might change the design of the SPAT.

(2)	Constructs	Humanness (DP1), Conversational Instruction (DP2), Dynamic Scaffolding (DP3), Engaging Dialog Flow (DP4), student engagement (for measuring the learning process DV), learning outcomes (for measuring outcome-related DVs such as problem-solving skills)
(3)	Principles of form and function	Part I: SPAT DPs DP1 – Humanness DP2 – Conversational instruction DP3 – Dynamic Scaffolding DP4 – Engaging Dialog Flow DP5 – Accessibility Part II: SPAT design process model 
(4)	Artifact mutability	This dissertation shows how SPAT can be adapted for different learning environments (online and blended learning environments) and learning outcomes (problem-solving

		skills, cognitive learning outcomes, etc.) in different experimental settings (laboratory and field settings). Suggestions for improvement of the DPs and the process: • DPs should be tested in isolation rather than mixing some of them together. • Platform should include a more detailed monitoring of students' learning processes with the help of a learning analytics approach. This can, for example, be shown in the dashboard of the platform, where educators see how many students answered how many question correctly/incorrectly. • More hints on how the interaction model can be created should be given when educators need to create SPAT for their learning environments. • More SPAT types should be included, so that educators can create the SPAT that is best suited to them.
(5)	Testable propositions	The theory offers testable predictions about the influence of the different DPs on students' learning processes and outcomes that are tested in different experiments. The DPs should also hold for other learning environment characteristics that are not tested. Moreover, the platform should help educators at different levels for different subjects. Specifically, if an educator follows the process and uses the platform, they will be able to better support students in the learning process compared to more traditional learning technologies in online and blended learning environments.
(6)	Justificatory knowledge	SPAT is able to trigger interactive learning behavior. According to the constructivist learning paradigm and the ICAP-framework in particular, students' interactive learning behavior leads to greater levels of learning success compared to constructive, active, and passive learning behavior. Based on that, all studies are embedded in a TML framework and all my DPs are derived from three different theoretical perspectives (education, psychology, information systems) (Vygotsky 1978).
(7)	Principle of implementation	The process as well as the created SPAT platform aid educators in creating properly designed SPAT and can thus be seen as a bridge to implementing the DPs.

(8)	Expository instantiation	The SPAT DPs were instantiated several times in the experiments where I created real SPAT. Moreover, the process and platform was instantiated as a realistic artifact that educators can use to create their own SPAT.
-----	--------------------------	---

Table 45. Design Theory for Providing Individual Support in Online and Blended Learning Environments

Source: Own illustration

7.1.4.1 Purpose and Scope

In this dissertation, I formulated DPs for SPAT in learning environments and developed a platform for educators to create and embed SPAT by themselves in order to address the problem of individual support in online and blended learning environments. The SPAT DP and the platform are mainly aimed at educators at different educational levels and dealing with different subjects. The platform is based on a self-developed process model and is influenced by the DPs. It considers a number of parameters that change the final result. The platform can be used by educators to create SPAT for their students. This SPAT can assist educators in providing individual support.

7.1.4.2 Constructs

Following the principle of persimony from Weber et al. (2012), the theory comprises only a few principles such as humanness, conversational instruction, dynamic scaffolding, and an engaging dialog flow. Humanness describes the level of anthropomorphism of a computer system (Doyle et al. 2019). It describes the extent to which human-like cues such as language style and name can positively influence the quality of interaction (Doyle et al. 2019). Conversational instruction describes how an educator transfers the content to the students. Dynamic scaffolding is the individual support given to a student by an instructor throughout the learning process. This support is specifically tailored to each student (Molenaar et al. 2012). An engaging dialog flow keeps the user motivated to continue the conversation (McTear et al. 2016). Learning process is a process that people pass through in order to acquire new knowledge and skills and ultimately influence their attitudes, decisions, and actions (Gupta and Bostrom 2009). Learning outcomes are statements that describe the knowledge, skills, and values as a whole to reflect how people really use knowledge (Gupta and Bostrom 2009).

7.1.4.3 Principles of Form and Function

In Section 4.2, I first proposed principles of form and function that I derived from theory and practice, and I then tested and further refined these DPs. Properly designed SPAT for education should show high levels of humanness, conversational instruction, and dynamic scaffolding, should use an engaging dialog flow, and should be accessible for educators and students. To address the latter point, I produced a process model to help educators to create SPAT according to this

DPs. Moreover, I created a platform that guides non-tech-savvy educators through the process model to develop SPAT for themselves.

7.1.4.4 Artifact Mutability

This dissertation shows how SPAT can be adapted for different learning environments (online and blended learning environments) and learning outcomes (problem-solving skills, cognitive learning outcomes, etc.) in different experimental settings (laboratory and field settings). There are also several suggestions for improving the developed DPs and the platform in future. First, the DPs should be tested individually to verify their effectiveness. Second, the process model should include the monitoring of students' learning processes by using learning analytics. This could be shown in the dashboard of the platform—for example, educators can see how many questions students answered correctly and where their main knowledge gaps are. Moreover, hints on how the interaction model can be created should be given when educators have to create SPAT for their learning environments. More SPAT types should be included, so that educators can create the SPAT that suits them best.

7.1.4.5 Testable Propositions

The theory proposed in this dissertation offers testable predictions about the influence of the different DPs on students' learning processes and outcomes tested in different experimental settings. The DPs should also hold for learning environments that were not tested in this dissertation. Moreover, the platform should help educators at different levels for different subjects. Specifically, if an educator follows the process and uses the platform, they will be able to better support their students in their learning processes compared to educators using other learning technologies in online and blended learning environments.

7.1.4.6 Justificatory knowledge

The kernel theories behind the derivation of DPs and the process model are grounded in a constructivist view of learning. The kernel theories include the ICAP-framework, the scaffolding theory, the computer as social actors paradigm, and learning engagement theory. Students' interactive learning behavior leads to greater levels of learning success compared to constructive, active, and passive learning behavior (ICAP-framework, Chi and Wylie 2014). Moreover, dynamic scaffolding builds students' knowledge leading to better learning outcomes (Molenaar et al. 2012). Furthermore, humans instinctively apply the social heuristics they use for human interactions when interacting with computers if the computer displays human-like social attributes (Nass et al. 1997; Reeves and Nass 1996). Furthermore, higher student engagement leads to better learning outcomes (Halverson and Graham 2019).

7.1.4.7 Principle of implementation

The SPAT, as well as the created SPAT platform, helps educators to instantiate the proposed DPs. The step-by-step guidelines helps educators to work systematically through the process and include all the individual parameters that are necessary to create an individualized SPAT for their online and blended learning environments.

7.1.4.8 Expository instantiation

The SPAT DPs were instantiated several times in the experiments in which I created real SPAT. Moreover, the process and platform were implemented in a realistic artifact that should help to evaluate the process model.

7.2 Practical Contributions of this Dissertation

The practical contributions of this dissertation are strongly related to the previously summarized theoretical contributions and can largely be divided into two main contributions.

7.2.1 Design Principles for Educators

The first practical contribution relates to the empirically tested DPs of RQ1-3. The DPs and the experimental implementations provide educators with evidence that these kinds of SPAT enable better learning outcomes and provide explicit guidance on how to translate abstract DPs such as humanness, conversational instruction, dynamic scaffolding, and an engaging dialog flow into precise DFs. They also provide explicit guidance on how to create and implement SPAT in online and blended learning environments.

In Section 5.2, I showed how a conversational instruction logic can be implemented by creating a dialog between SPAT and students using a proactive and reactive logic. With the proactive logic, the SPAT proactively guided students through the task by using five different problem-solving activities adapted from Kim and Hannafin (2011). With the reactive logic, the SPAT was able to provide feedback and react to students' questions that arose during the different problem-solving steps. Moreover, I showed how to implement humanness by adapting SPAT utterances to human tutorial dialogs (see Section 5.1). In Section 5.3, I demonstrated how to create and technically implement a dynamically scaffolding dialog between SPAT and students by making a division between the main and sub dialog. Whenever the student did not know the answer, the SPAT changed to the sub dialog and tried to bring students back on track. The field experiments also showed how to implement SPAT in real-life learning environments by using SPAT as individual support for homework assignments (see Section 5.2). In Section 5.3, I demonstrated how SPAT interactions during online video lectures have to be created in order to improve student engagement while watching the videos.

These implementations provide guidance for educators who want to create and embed SPAT in online and blended learning environments. These implementations might help SPAT designers to break down DPs for their own context by mimicking the implementation in the paper. As such, the created SPAT and implementations should be generalizable to other online and blended learning environments to address the problem of individual support in these environments.

7.2.2 Smart Personal Assistant Technology Platform

While the DPs are a first orientation for educators, the SPAT platform additionally helps to create SPAT in a very efficient and effective way. The platform is divided into a preceding process model and the web user interface itself. The process model provides educators with explicit guidance on how to integrate SPAT into their didactical concept and how to give students access to this technology. The web user interface is directly based on the process model by including every step to create and embed SPAT in online and blended learning environments. If educators want to integrate digital learning technologies into their learning environments, they often depend on the technical know-how of the software developers. This leads to educators not accepting this technology, despite using it, and they seldom feel empowered to provide the technology to their students. My platform will give educators a helping hand in creating and implementing SPAT for themselves in order to offer individual support for students in online and blended learning environments. This platform will also enable educators to access existing SPAT from other educators and analyze the learning progress of students. Consequently, the adaption barriers of educators will be lowered down and educators will be empowered to use SPAT for themselves.

7.3 Directions for Future Research

“With the hype around bots, everyone is running around like headless chickens to build bots that will solve every existing problem in the world. (Bottr 2016)”

In this section of my dissertation, I suggest areas and directions for future research activities. While the limitations and directly connected future research possibilities were discussed in the specific studies, I will now focus on overarching themes and research questions for future research. Figure 49 presents future research areas and their relationship. It is practical to look at future research areas from three perspectives: educational, conversation design, and technical perspective. To unleash the full potential of SPAT in blended and online learning environments, the pertinent research questions must be answered in all three areas in the future.

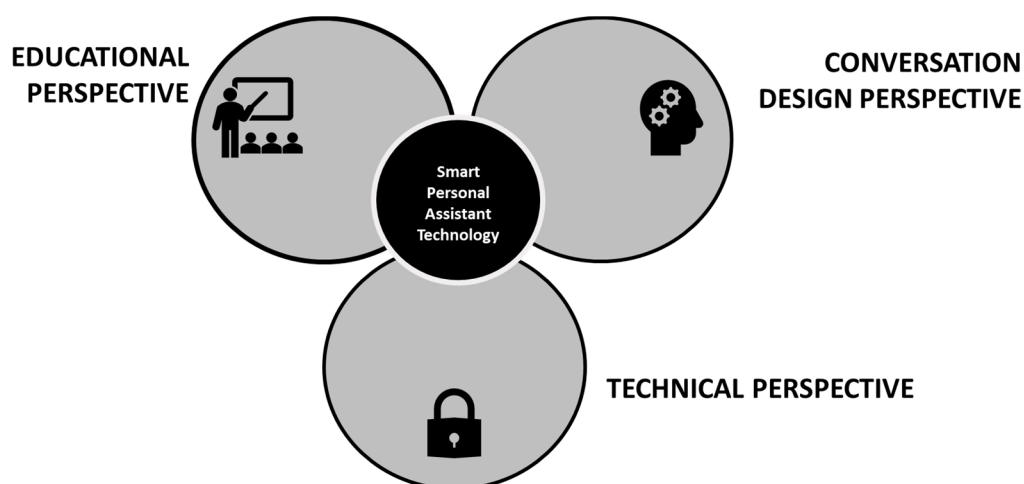


Figure 49. SPAT Future Areas of Research

Source: Own illustration

7.3.1 Educational Perspective

This perspective of this research is connected to the possible new use cases for SPAT in education and to the way SPAT can influence learning processes. During the dissertation development process, some potentially interesting areas of application for SPAT were identified. The most promising use environments might be SPAT for collaborative learning (Graesser et al. 2017a), embodied SPAT for human-like tutoring in online learning (Fu et al. 2019), and SPAT for improving metacognitive skills (Matsuda et al. 2020).

Regarding collaborative learning, in Section 5.1 I showed how SPAT can help groups of three to increase their task performance in a laboratory setting and also identified first interaction patterns that differ to those associated with human tutors. This study is a starting point for a new line of research that raises new questions. Automatic analysis of collaborative processes is an advance in the field of NLP that has value for real time assessment during collaborative learning, for dynamically triggering supportive interventions in the midst of collaborative-learning sessions, and for facilitating efficient analysis of collaborative-learning processes (Hernández-Sellés et al. 2019). Research in the field of computer-supported collaborative learning has revealed that unsolicited conversational agent interventions can intensify the knowledge exchange among learning partners and increase students' explicit reasoning and participation levels (Adamson et al. 2013). Agent-based supportive mechanisms can positively impact the quality and conceptual depth of students' conversations and, consequently, the learning outcomes (Tegos and Demetriadis 2017). Particularly in the context of online learning, SPAT can be very useful for providing learning support during the course to compensate for the often insufficient individual support of educators, which constitutes one of the key factors negatively affecting retention rates (Hone and El Said 2016). While the potential in this area has already been identified by initial studies, many questions remain. Possible research questions arising from this dissertation include the following:

- What role should SPAT play in collaborative learning settings?
- What are possible use cases for SPAT in collaborative learning settings?
- How do students perceive SPAT as collaborative learning support?
- How do interactions with SPAT influence students' learning processes and long-term skill development?

The second promising usage environment for SPAT is embodied SPAT for human-like tutoring in online learning (Fu et al. 2019). The area of multimodal conversational AI promises to provide visual, audio, and other sensor inputs simultaneously in order to enable capabilities such as multi-context conversations. These technological advances open another interesting field of research, which can be described as multimodal SPAT in online courses. Technology leaders such as Nvidia have already presented SPAT that is able to react with gestures and facial expressions in real time, giving students the feeling of talking to a real person on the screen (Llopert 2020). In Section 5.3, I integrated a SPAT in an online course video and tested its effectiveness within a laboratory experiment. The results showed that learning engagement can be increased. These initial findings raise new questions such as how multimodal, multi-sensor-based SPAT can mimic real human tutor interactions thereby addressing the problem of individual support in online and blended learning environments. In the not too distant future, it is even conceivable that this multimodal SPAT can take over the role of the teaching assistant, interrupting instruction phases in real time and then responding to individual students' questions. Until then, there are still some questions to be answered. Possible research questions that connect to this dissertation include the following:

- How do students interact with multi-modal, multi-sensor based SPAT compared to voice-only SPAT?
- How do students interact with multi-modal, multi-sensor based SPAT compared to human tutors?
- Does multi-modal, multi-sensor based SPAT have an effect on long-term skill development compared to voice-only SPAT and also compared to video-only in online learning environments?
- How can multi-modal, multi-sensor based SPAT be designed as human-like tutors in online and blended learning environments?

Regarding SPAT for improving metacognitive skills, many recent studies have tried to research this area and test the first effects (Wambsganss et al. 2020). The Organization for Economic Co-operation and Development (OECD) highlights the importance of metacognitive skills in carrying out processes and enabling the use of one's own knowledge in a responsible way to achieve a goal. Metacognitive skills include learning-to-learn skills and the ability to recognize one's knowledge, skills, attitudes, and values (Dede 2010). For example, SPAT can prompt students' self-regulated learning aspects. The SPAT's responses can

be specifically designed to stimulate students' metacognitive skills. For example, during the conversation, students can be asked such things as whether they plan their learning, use specific learning strategies, or evaluate their learning performance. (Delen et al. 2014). In Section 5.2, I investigated the impact of SPAT conversations on problem-solving skills. While this is a first step into this research area, less is known about how SPAT-based conversations affect the development of metacognitive skills such as creativity, critical thinking, and self-regulated learning. Possible research questions in this line of research include the following:

- How can SPAT be used to improve students' creativity processes?
- How can SPAT be used to improve students' critical thinking skills?
- How can SPAT be used to improve students' self-regulated learning skills?
- How can SPAT be designed to improve students' metacognitive skills?

7.3.2 Conversation Design Perspective

Creating meaningful conversations with SPAT that improve students' learning processes and learning outcomes requires a breadth of learning and design expertise. The role of a SPAT conversation designer for online and blended learning environments is similar to that of an architect, mapping out what students can do in a space while considering both the students' needs and the technological constraints. They curate the conversation and define the flow and its underlying logic in a detailed design specification that represents the complete UE. The research area of conversational scripting in education emphasizes how dialogues can be pre-scripted in order to facilitate learning (Fischer et al. 2013). The more a computer-supported collaborative learning practice differs from traditional teaching and learning experiences, the more difficult it may be for students to collaborate efficiently. Students may not have developed adequate knowledge that guides them in collaborating in these settings. There is a substantial number of empirical studies on the effects of external collaboration scripts on learning processes and outcomes (Fischer et al. 2013; O'shea 2014; Rummel and Spada 2005). More and more studies have based their conversation designs on such scripts. However, recent developments in NLP allow the design of more open dialogues that focus on a more flexible interaction flow rather than a pre-defined script. It is still not clear which use cases fit best with the use of SPAT or how to best create interaction flows and integrate these dialogues successfully into existing online and blended course formats. Particularly in the context of education, it is not yet certain which steps need to be taken to create efficient interaction dialogues in a systematic way.

Measuring whether a use case is appropriate for the use of SPAT is the first step in a conversation design process. There is still too little about the kind of learning contexts and learning goals that fit the use of SPAT. For example, conversation

dialog may work well for receiving new information, but it feels inefficient for actively browsing new information on the internet. Relevant research questions in this area include the following:

- What are the relevant decision criteria for evaluating use cases for conversation design in education?
- Which learning goals fit best to the use of conversation design?
- Which learning activities fit best to the use of conversation design?

Future research should also consider the interaction logic in detail. For example, in Section 5.3, I used a scaffolding mechanism in my SPAT to offer dynamic support for students. It would be helpful to determine which other mechanisms and strategies should be incorporated into SPAT in order to impact the learning process and outcomes. Possible research questions in this field include the following:

- How can an instructional conversation flow be created?
- How should sample dialogs best be written?
- How can these SPAT dialogues be successfully integrated into existing online and blended learning courses to provide students with access to the SPAT?

7.3.3 Technical Perspective

Major tech companies such as Amazon and Google have invested heavily in the technologies required to develop sophisticated systems capable of interacting with users using a natural, conversational style, such as AI, particularly deep learning and NLP. There are several reasons why SPAT can be seen as a different class of system compared to older generations of intelligent tutoring systems. There are various technological drivers that have facilitated the development and deployment of SPAT, including: a) advances in deep learning and deep neural networks; b) greater computing processing power to support the massive parallel computations required to run deep neural networks; c) the availability of vast amounts of data that enables SPAT to learn and become increasingly more intelligent; d) increased connectivity, allowing users to connect their smart devices to vast cloud-based resources; and e) advances in speech recognition and NLP technologies, mainly as a result of the application of deep neural networks (McTear 2017). The relevant research areas are deep neural networks, research into new corpora building in the area of skill development, and advancements in the area of speech recognition and research around learning analytics (e.g., how to use the generated data in a meaningful way?).

Future research should try to fuse different kinds of deep neural networks in order to provide the best experience for students. Possible research questions in this area include the following:

- How can deep neural network algorithms be combined to best detect students' user intents?
- How can reinforcement learning mechanisms in task-oriented dialogue agents be improved so that the systems can learn by themselves?

Moreover, research should focus on building new corpora in the area of skill development, so that SPAT can be developed accordingly and perhaps provided as a cloud-based resource. This would help future researchers and practitioners in building better systems. Relevant research questions in this field include:

- What must a corpora look like for the development of creative skills?
- What must a corpora look like for the development of critical thinking skills?
- What must a corpora look like for the development of self-regulated learning skills?
- How can educators use these corpora to create SPAT by themselves?

Moreover, research on how speech recognition can be pushed to a level where it sounds less artificial and feels similar to real human tutor dialogs is another interesting area of inquiry. For example, Parviainen and Sondergaard (2020) have investigated the influence of the whispering mode of Amazon's Alexa on user perceptions. This is an interesting option, as students may wish to use SPAT in public places. Relevant research questions in this research direction include the following:

- How do different voice characteristics influence students' perception of SPAT?
- Does the voice influence students' learning engagement and resulting learning outcome?
- How can the voice be made even more human?

The subject of learning analytics is also a very promising topic for future research. When students interact with SPAT, they create a lot of qualitative data. It is yet not clear how these data can be used to inform educators' and students' decisions. Relevant research questions in this field include the following:

- How can conversational log data be used to inform educators' decisions?
- How can conversational log data be used to inform students' decisions?
- How can conversational log data automatically be extracted with techniques such as text mining, etc.?

8 References

- Abdul-Kader, Sameera A., and Woods, J. 2015. "Survey on chatbot design techniques in speech conversation systems," *Int. Journal on Advanced Computer Science Applied* (6:7), pp. 72-80.
- Abu Shawar, B., and Atwell, E. 2007. "Chatbots: Are they really useful?" *LDV Forum* (22:1), pp. 29-49.
- Adamson, D., Ashe, C., Jang, H., Yaron, D., and Rosé, C. P. 2013. "Intensification of Group Knowledge Exchange with Academically Productive Talk Agents," in *Intensification of Group Knowledge Exchange with Academically Productive Talk Agents*, The Computer Supported Collaborative Learning (CSCL) Conference 2013 (ed.).
- Adamson, D., Dyke, G., Jang, H., and Rosé, C. P. 2014. "Towards an agile approach to adapting dynamic collaboration support to student needs," *International Journal of Artificial Intelligence in Education* (24:1), pp. 92-124.
- Albacete, P., Jordan, P., Lusetich, D., Chounta, I. A., Katz, S., and McLaren, B. M. 2018. "Providing Proactive Scaffolding During Tutorial Dialogue Using Guidance from Student Model Predictions," in *Artificial Intelligence in Education*, C. Penstein Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren and B. Du Boulay (eds.), Cham: Springer International Publishing, pp. 20-25.
- Alepis, E., and Virvou, M. 2011. "Automatic generation of emotions in tutoring agents for affective e-learning in medical education," *Expert Systems with Applications* (38:8), pp. 9840-9847.
- Allison, D. 2012. "Chatbots in the library: Is it time?" *Library Hi Tech* (30:1), pp. 95-107.
- Amazon 2019. *New Amazon Alexa Skill Kit 2.0*. <https://developer.amazon.com/blogs/alexa/post/1a4e8b01-663d-4680-8efd-c28e96e31655/now-available-version-2-of-the-ask-software-development-kit-for-java>. Accessed 20 November 2018.
- Ammar, M. B., Neji, M., Alimi, A. M., and Gouardères, G. 2010. "The affective tutoring system," *Expert Systems with Applications* (37:4), pp. 3013-3023.
- Anderson, L. W., and Krathwohl, D. R. 2001. "A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives," *New York. Longman*. (9:2), pp. 137-175.
- Anderson, T., and Kanuka, H. 1999. "Using constructivism in technology-mediated learning: Constructing order out of the chaos in the literature," 1524-6345.

- Angeli, A. de, and Brahnam, S. 2008. "I hate you! Disinhibition with virtual partners," *Interacting with Computers* (20:3), pp. 302-310.
- Anson, R., Bostrom, R., and Wynne, B. 1995. "An Experiment Assessing Group Support System and Facilitator Effects on Meeting Outcomes," *Management Science* (41:2), pp. 189-208.
- Applebee, A. N., and Langer, J. A. 1983. "Instructional scaffolding: Reading and writing as natural language activities," *Language arts* (60:2), pp. 168-175.
- Arend, B. 2018. "Investigating Siri as a virtual assistant in a learning context," in *Proceedings of 12th annual International Technology, Education and Development Conference*, Seville, Spain, pp. 1-10.
- Arksey, H., and O'Malley, L. 2005. "Scoping studies: Towards a methodological framework," *International journal of social research methodology* (8:1), pp. 19-32.
- Atkinson, J. M., and Heritage, J. 1984. *Structures of social action*, Cambridge University Press.
- AXA Group Operations 2018. *NLP.js Framework*. <https://github.com/axa-group/nlp.js>.
- Ayedoun, E., Hayashi, Y., and Seta, K. 2015. "A Conversational Agent to Encourage Willingness to Communicate in the Context of English as a Foreign Language," *Procedia Computer Science* (60:1), pp. 1433-1442.
- Azevedo, R., Cromley, J. G., and Seibert, D. 2004. "Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia?" *Contemporary Educational Psychology* (29:3), pp. 344-370.
- Backhaus, K., Erichson, B., Plinke, W., and Weiber, R. 2016. *Multivariate analysemethoden*, Springer.
- Baker, R. S. 2016. "Stupid tutoring systems, intelligent humans," *International Journal of Artificial Intelligence in Education* (26:2), pp. 600-614.
- Bakki, A., Oubahssi, L., George, S., and Cherkaoui, C. 2019. "MOOCAT: A visual authoring tool in the cMOOC context," *Education and Information Technologies* (24:2), pp. 1185-1209.
- Baldwin, S. A., Stice, E., and Rohde, P. 2008. "Statistical analysis of group-administered intervention data: reanalysis of two randomized trials," *Psychotherapy research : journal of the Society for Psychotherapy Research* (18:4), pp. 365-376.

- Bannert, M., and Mengelkamp, C. 2008. "Assessment of metacognitive skills by means of instruction to think aloud and reflect when prompted. Does the verbalisation method affect learning?" *Metacognition and Learning* (3:1), pp. 39-58.
- Bannert, M., Molenaar, I., and Azevedo, R., Järvelä, S., Gasevic D. (eds.) 2017. *Relevance of learning analytics to measure and support students' learning in adaptive educational technologies*.
- Barbhuiya, R., Karim, K. M., and Suraiya, J. 2011. "Design specifications for a generic Intelligent Tutoring System," in *Proceedings of the International Conference on e-Learning, e-Business, Enterprise Information Systems, and e-Government (EEE)*, Las Vegas, USA: The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
- Barricelli, B. R., Cassano, F., Fogli, D., and Piccinno, A. 2019. "End-user development, end-user programming and end-user software engineering: A systematic mapping study," *Journal of Systems and Software* (149), pp. 101-137.
- Belanger, F. 2012. "Theorizing in information systems research using focus groups," *Australasian Journal of Information Systems* (17:2), pp. 109-135.
- Ben Mimoun, Mohammed Slim, and Poncin, I. 2015. "A valued agent: How ECAs affect website customers' satisfaction and behaviors," *Journal of Retailing and Consumer Services* (26), pp. 70-82.
- Benbasat, I., and Zmud, R. W. 2003. "The identity crisis within the IS discipline: Defining and communicating the discipline's core properties," *MIS quarterly*, pp. 183-194.
- Berge, T. ten, and van Hezewijk, R. 1999. "Procedural and declarative knowledge: An evolutionary perspective," *Theory & Psychology* (9:5), pp. 605-624.
- Berry, D. C., Butler, L. T., and Rosis, F. de 2005. "Evaluating a realistic agent in an advice-giving task," *International Journal of Human-Computer Studies* (63:3), pp. 304-327.
- Bhattacharjee, A. 2012. *Social science research: Principles, methods, and practices*.
- Bickmore, Timothy, and Cassell, J. 1999. "Small talk and conversational storytelling in embodied conversational interface agents," in *Small talk and conversational storytelling in embodied conversational interface agents*.

- Bickmore, T. W., Mitchell, S. E., Jack, B. W., Paasche-Orlow, M. K., Pfeifer, L. M., and Odonnell, J. 2010. "Response to a Relational Agent by Hospital Patients with Depressive Symptoms," *Interacting with Computers* (22:4), pp. 289-298.
- Bickmore, T. W., Schulman, D., and Sidner, C. 2013. "Automated interventions for multiple health behaviors using conversational agents," *Patient Education and Counseling* (92:2), pp. 142-148.
- Bitzer, P., and Janson, A. "Towards a Holistic Understanding of Technology Mediated Learning Sciences," in *European Conference on Information Systems 2014 (ECIS)*. Tel Aviv, Israel.
- Blake, W. 2007. *Blake: the complete poems*, Pearson Education.
- Bonate, P. L. 2000. *Analysis of pretest-posttest designs*, CRC Press.
- Bonwell, C. C., and Eison, J. A. 1991. *Active Learning: Creating Excitement in the Classroom*. ASHE-ERIC Higher Education Reports, ERIC.
- Botsman, R. 2017. "Big data meets Big Brother as China moves to rate its citizens," *Wired UK* (21), pp. 1-11.
- Bottr, A. G. 2016. *Why we are still a long way from a "killer" bot*. <https://venturebeat.com/2016/09/19/why-we-are-still-a-long-way-from-a-killer-chatbot/>. Accessed 14 November 2020.
- Boucsein, W. 2013. *Elektrodermale Aktivität: Grundlagen, Methoden und Anwendungen*, Springer-Verlag.
- Brackett, M. A., and Katulak, N. A. 2007. "Emotional intelligence in the classroom: Skill-based training for teachers and students," *Applying emotional intelligence: A practitioner's guide*, pp. 1-27.
- Brahnam, S., and Angeli, A. de 2012. "Gender affordances of conversational agents," *Interacting with Computers* (24:3), pp. 139-153.
- Briggs, R. O. 2006. "On theory-driven design and deployment of collaboration systems," *International Journal of Human-Computer Studies* (64:7), pp. 573-582.
- Briggs, R. O., Kolfshoten, G. L., Vreede, G.-J. de, Lukosch, S., and Albrecht, C. C. 2013. "Facilitator-in-a-box: process support applications to help practitioners realize the potential of collaboration technology," *Journal of Management Information Systems* (29:4), pp. 159-194.
- Briggs, R. O., Vreede, G.-J. de, and Nunamaker Jr, J. F. 2003. "Collaboration engineering with ThinkLets to pursue sustained success with group support systems," *Journal of Management Information Systems* (19:4), pp. 31-64.

- Brinkkemper, S. 1996. "Method engineering: engineering of information systems development methods and tools," *Information and Software Technology* (38:4), pp. 275-280.
- Brinton, C. G., Rill, R., Ha, S., Chiang, M., Smith, R., and Ju, W. 2015. "Individualization for Education at Scale: MIIC Design and Preliminary Evaluation," *IEEE Transactions on Learning Technologies* (8:1), pp. 136-148.
- Britz, D. 2016. *Deep Learning for Chatbots, Part 1 – Introduction*. <http://www.wildml.com/2016/04/deep-learning-for-chatbots-part-1-introduction/>. Accessed 5 January 2018.
- Brush, T. A., and Saye, J. W. 2002. "A summary of research exploring hard and soft scaffolding for teachers and students using a multimedia supported learning environment," *The Journal of Interactive Online Learning* (1:2), pp. 1-12.
- Bull, K. S., Shuler, P., Overton, R., Kimball, S., Boykin, C., and Griffin, J. 1999. "Processes for Developing Scaffolding in a Computer Mediated Learning Environment," *Rural Special Education for the New Millennium*. (10:1), pp. 18-45.
- Canbek, N. G., and Mutlu, M. E. 2016. "On the track of artificial intelligence: Learning with intelligent personal assistants," *Journal of Human Sciences* (13:1), pp. 592-601.
- Chad Lane, H., Core, M. G., Hays, M. J., Auerbach, D., and Rosenberg, M. 2015. "Situated Pedagogical Authoring: Authoring Intelligent Tutors from a Student's Perspective," in *Artificial Intelligence in Education*, C. Conati, N. Heffernan, A. Mitrovic and M. F. Verdejo (eds.), Cham: Springer International Publishing, pp. 195-204.
- Chaiklin, S. 2003. "The zone of proximal development in Vygotsky's analysis of learning and instruction," *Vygotsky's educational theory in cultural context* (1:2), pp. 39-64.
- Cheepen, C. 1988. *The predictability of informal conversation*, Pinter Pub Ltd.
- Chen, C.-M. 2008. "Intelligent web-based learning system with personalized learning path guidance," *Computers & Education* (51:2), pp. 787-814.
- Cheung, A. C. K., and Slavin, R. E. 2013. "The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis," *Educational Research Review* (9:1), pp. 88-113.
- Chi, M. T. H., and Wylie, R. 2014. "The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes," *Educational Psychologist* (49:4), pp. 219-243.

- Christensen, G., Steinmetz, A., Alcorn, B., Bennett, A., Woods, D., and Emanuel, E. 2013. "The MOOC phenomenon: who takes massive open online courses and why?" *SSRN* (2:24), pp. 1-25.
- Chung, J. 1991. "Collaborative learning strategies: The design of instructional environments for the emerging new school," *Educational Technology* (31:12), pp. 15-22.
- Cognition and Technology Group 1991. "Technology and the design of generative learning environments," *Educational Technology*, pp. 34-40.
- Cohen, J. 1960. "A coefficient of agreement for nominal scales," *Educational and psychological measurement* (20:1), pp. 37-46.
- Cohen, J. 1973. "Eta-squared and partial eta-squared in fixed factor ANOVA designs," *Educational and psychological measurement* (33:1), pp. 107-112.
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences 2nd ed.*, Erlbaum Associates, Hillsdale.
- Cohen, J. 1992. "A power primer," *Psychological Bulletin* (112:1), pp. 155-159.
- Cohn, M. 2004. *User stories applied: For agile software development*, Addison-Wesley Professional.
- Cooper, H. M. 1988. "Organizing knowledge syntheses: A taxonomy of literature reviews," *Knowledge, Technology & Policy* (1:1), pp. 104-126.
- Copley, J. 1992. "The integration of teacher education and technology: a constructivist model," in *Technology and Teacher Education*, D. Carey, R. Carey, D. Willis and Willis J. (eds.), Charlottesville.
- Corley, K. G., and Gioia, D. A. 2011. "Building theory about theory building: what constitutes a theoretical contribution?" *Academy of management review* (36:1), pp. 12-32.
- Corti, K., and Gillespie, A. 2016. "Co-constructing intersubjectivity with artificial conversational agents: People are more likely to initiate repairs of misunderstandings with agents represented as human," *Computers in Human Behavior* (58), pp. 431-442.
- Cortright, R. N., Collins, H. L., and DiCarlo, S. E. 2005. "Peer instruction enhanced meaningful learning: ability to solve novel problems," *Advances in physiology education* (29:2), pp. 107-111.
- Cremonesi, P., Ricci, F., Berkovsky, S., and Tuzhilin, A. (eds.) 2017. *RecSys '17: Proceedings of the Eleventh ACM Conference on Recommender Systems, August 27-31, 2017, Como, Italy, Como, Italy, 27.08.2017 - 31.08.2017*, New York: Association for Computing Machinery.

- Creswell, J. W. 2013. *Research design: Qualitative, quantitative, and mixed methods approaches*, Sage Publications.
- Creswell, J. W. 2014. "The selection of a research approach," *Research design: Qualitative, quantitative, and mixed methods approaches*, pp. 3-24.
- Crockett, K., Latham, A., and Whitton, N. 2017. "On predicting learning styles in conversational intelligent tutoring systems using fuzzy decision trees," *International Journal of Human-Computer Studies* (97), pp. 98-115.
- Cukurova, M. 2019. "Learning analytics as AI extenders in education: Multimodal machine learning versus multimodal learning analytics," in *Learning analytics as AI extenders in education: Multimodal machine learning versus multimodal learning analytics*, Proceedings of the Artificial Intelligence and Adaptive Education Conference (ed.), Chicago, USA.
- Damacharla, P., Dhakal, P., Stumbo, S., Javaid, A. Y., Ganapathy, S., Malek, D. A., Hodge, D. C., and Devabhaktuni, V. 2019. "Effects of Voice-Based Synthetic Assistant on Performance of Emergency Care Provider in Training," *International Journal of Artificial Intelligence in Education* (29:1), pp. 122-143.
- Dede, C. 2010. "Comparing frameworks for 21st century skills," *21st century skills: Rethinking how students learn* (20:2010), pp. 51-76.
- Delen, E., Liew, J., and Willson, V. 2014. "Effects of interactivity and instructional scaffolding on learning: Self-regulation in online video-based environments," *Computers & Education* (78), pp. 312-320.
- Dennis, A. R., and Valacich, J. S. 2001. "Conducting experimental research in information systems," *Communications of the Association for Information Systems* (7:1), p. 5.
- DeSanctis, G., and Poole, M. S. 1994. "Capturing the complexity in advanced technology use: Adaptive structuration theory," *Organization science* (5:2), pp. 121-147.
- Devlin, W. J., and Bokulich, A. 2015. *Kuhn's Structure of Scientific Revolutions-50 years on*, Springer.
- Dillenbourg, P., Lemaignan, S., Sangin, M., Nova, N., and Molinari, G. 2016. "The symmetry of partner modelling," *International Journal of Computer-Supported Collaborative Learning* (11:2), pp. 227-253.
- Dimitrov, D. M., and Rumrill, P. D. 2003. "Pretest-posttest designs and measurement of change," *Work* (20:2), pp. 159-165.
- Dimoka, A., Davis, F. D., Gupta, A., Pavlou, P. A., Banker, R. D., Dennis, A. R., Ischebeck, A., Müller-Putz, G., Benbasat, I., and Gefen, D. 2012. "On the use

- of neurophysiological tools in IS research: Developing a research agenda for NeuroIS," *MIS quarterly*, pp. 679-702.
- D'Mello, S. K., Graesser, A., and King, B. 2010. "Toward Spoken Human-Computer Tutorial Dialogues," *Human-Computer Interaction* (25:4), pp. 289-323.
- Döring, N., and Bortz, J. 2016. "Forschungsmethoden und Evaluation," *Wiesbaden: Springerverlag*.
- Dousay, T. A., and Hall, C. 2018. "Alexa, tell me about using a virtual assistant in the classroom," in *Proceedings of EdMedia: World Conference on Educational Media and Technology*, Amsterdam, Netherlands, Amsterdam, Netherlands: Association for the Advancement of Computing in Education (AACE), pp. 1-19.
- Doyle, P. R., Edwards, J., Dumbleton, O., Clark, L., and Cowan, B. R. 2019. "Mapping Perceptions of Humanness in Intelligent Personal Assistant Interaction," in *Proceedings of the 21st International 2019*, pp. 1-12.
- Drew, P. 2005. "Conversation analysis," *Handbook of language and social interaction* (2:3), pp. 71-102.
- Duffy, M. C., and Azevedo, R. 2015. "Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system," *Computers in Human Behavior* (52), pp. 338-348.
- Dumford, A. D., and Miller, A. L. 2018. "Online learning in higher education: exploring advantages and disadvantages for engagement," *Journal of Computing in Higher Education* (30:3), pp. 452-465.
- Dutta, D. 2017. "Developing an Intelligent Chatbot Tool to assist high school students for learning general knowledge subjects," Georgia Institute of Technology, Atlanta.
- Easwara Moorthy, A., and Vu, K.-P. L. 2015. "Privacy Concerns for Use of Voice Activated Personal Assistant in the Public Space," *International Journal of Human-Computer Interaction* (31:4), pp. 307-335.
- Educause 2019. *2019 Horizon Report*. <https://library.educause.edu/resources/2019/4/2019-horizon-report>. Accessed 14 November 2020.
- Ertmer, P. A., and Newby, T. J. 1993. "Behaviorism, Cognitivism, Constructivism: Comparing Critical Features from an Instructional Design Perspective," *Performance Improvement Quarterly* (6:4), pp. 50-72.
- Evans, C., and Gibbons, N. J. 2007. "The interactivity effect in multimedia learning," *Computers & Education* (49:4), pp. 1147-1160.

- Ferschke, O., Yang, D., Tomar, G., and Rosé, C. P. 2015. "Positive Impact of Collaborative Chat Participation in an edX MOOC," in *Artificial Intelligence in Education*, C. Conati, N. Heffernan, A. Mitrovic and M. F. Verdejo (eds.), Cham: Springer International Publishing, pp. 115-124.
- Fischer, F., Kollar, I., Stegmann, K., and Wecker, C. 2013. "Toward a script theory of guidance in computer-supported collaborative learning," *Educational Psychologist* (48:1), pp. 56-66.
- Fischer, G., Fogli, D., and Piccinno, A. 2017. "Revisiting and broadening the meta-design framework for end-user development," in *New perspectives in end-user development*, Springer, pp. 61-97.
- Flanagan, J. C. 1954. "The critical incident technique," *Psychological Bulletin* (51:4), p. 327.
- Fleiss, J. L., Levin, B., and Paik, M. C. 1981. "The measurement of interrater agreement," *Statistical methods for rates and proportions* (2:212-236), pp. 22-23.
- Foster, M. E., and Oberlander, J. 2010. "User preferences can drive facial expressions: Evaluating an embodied conversational agent in a recommender dialogue system," *User modeling and user-adapted interaction* (20:4), pp. 341-381.
- Fraenkel, J. R., Wallen, N. E., and Hyun, H. H. 2011. *How to design and evaluate research in education*, New York: McGraw-Hill Humanities/Social Sciences/Languages.
- Francis, A. L., and Oliver, J. 2018. "Psychophysiological measurement of affective responses during speech perception," *Hearing research* (369), pp. 103-119.
- French, A., Macedo, M., Poulsen, J., Waterson, T., and Yu, A. 2008. "Multivariate analysis of variance (MANOVA)," *San Francisco State University*.
- Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., and Sherlock, Z. 2017. "Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners," *Computers in Human Behavior* (75:10), pp. 461-468.
- Fu, Y., Wang, C., Fu, Y., Wang, Y.-X., Bai, C., Xue, X., and Jiang, Y.-G. 2019. "Embodied One-Shot Video Recognition," in *Proceedings of the 27th ACM International Conference on Multimedia*, L. Amsaleg, B. Huet, M. Larson, G. Gravier, H. Hung, C.-W. Ngo and W. Tsang Ooi (eds.), Nice France. 2019, New York, NY, USA: ACM, pp. 411-419.

- Gallagher, S. A., Stepien, W. J., and Rosenthal, H. 1992. "The effects of problem-based learning on problem solving," *Gifted Child Quarterly* (36:4), pp. 195-200.
- Gehl, R. W. 2014. "Teaching to the Turing Test with Cleverbot," *Transformations: The Journal of Inclusive Scholarship and Pedagogy* (24:1-2), pp. 56-66.
- Generalized Intelligent Framework for Tutoring (GIFT) Users Symposium (GIFTSym3)* 2015, Orlando, USA.
- Git Hub 2019. *NLP Framework BenchmarkING*. <https://github.com/axa-group/nlp.js/blob/master/docs/benchmarking.md>. Accessed 23 November 2020.
- Glaserfeld, E. v. 1987. "Constructivism," *The concise Corsini encyclopedia of psychology and behavioral science* (6), pp. 19-21.
- Goksel-Canbek, N., and Mutlu, M. E. 2016. "On the track of Artificial Intelligence: Learning with Intelligent Personal Assistants," *International Journal of Human Sciences* (13:1), p. 592.
- Goldkuhl, G. 2012. "Pragmatism vs interpretivism in qualitative information systems research," *European Journal of Information Systems* (21:2), pp. 135-146.
- Goleman, D., Boyatzis, R. E., and McKee, A. 2013. *Primal leadership: Unleashing the power of emotional intelligence*, Harvard Business Press.
- Goodwin, C. 1980. "Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning," *Sociological inquiry* (50:3-4), pp. 272-302.
- Goos, G., Hartmanis, J., van Leeuwen, J., Goettl, B. P., Halff, H. M., Redfield, C. L., and Shute, V. J. (eds.) 1998. *Intelligent Tutoring Systems*, Berlin, Heidelberg: Springer Berlin Heidelberg.
- Govindasamy, M. K. 2014. "Animated Pedagogical Agents: A Review of Agent Technology Software in Electronic Learning Environments," *Journal of Educational Multimedia and Hypermedia* (23:2), pp. 163-188.
- Graesser, A., Li, H., and Forsyth, C. 2014. "Learning by Communicating in Natural Language with Conversational Agents," *Grantee Submission*.
- Graesser, A., and McDaniel, B. 2017. "Conversational agents can provide formative assessment, constructive learning, and adaptive instruction," in *The future of assessment*, Routledge, pp. 85-112.
- Graesser, A. C., Cai, Z., Hu, X., Foltz, P. W., Greiff, S., and Shaffer, D. W. 2017a. "Assessment of collaborative problem solving," *Design Recommendations for Intelligent Tutoring Systems* (5:2), pp. 275-285.

- Graesser, A. C., Cai, Z., Morgan, B., and Wang, L. 2017b. "Assessment with computer agents that engage in conversational dialogues and trialogues with learners," *Computers in Human Behavior* (76), pp. 607-616.
- Graesser, A. C., Dowell, N., Hampton, A. J., Lippert, A. M., Li, H., and Shaffer, D. W. 2018a. "Building Intelligent Conversational Tutors and Mentors for Team Collaborative Problem Solving: Guidance from the 2015 Program for International Student Assessment," in *Building intelligent tutoring systems for teams: What matters*, J. Johnston, R. Sottolare, A. M. Sinatra and C. S. Burke (eds.), United Kingdom: Emerald Publishing, pp. 173-211.
- Graesser, A. C., Fiore, S. M., Greiff, S., Andrews-Todd, J., Foltz, P. W., and Hesse, F. W. 2018b. "Advancing the science of collaborative problem solving," *Psychological Science in the Public Interest* (19:2), pp. 59-92.
- Graesser, A. C., Hu, X., and Sottolare, R. 2018c. "Intelligent tutoring systems," in *International handbook of the learning sciences*, Routledge, pp. 246-255.
- Grant, R. H., Hewitt, T. L., Mason, M. J., Moore, R. J., and Winburn, K. A. 2019. *Cognitive intervention for voice recognition failure*, Google Patents: Google Patents.
- Gratch, J., Wang, N., Okhmatovskaia, A., Lamothe, F., Morales, M., van der Werf, R. J., and Morency, L.-P. 2007. "Can Virtual Humans Be More Engaging Than Real Ones?" in *Human-computer interaction: 12th international conference, HCI International 2007, Beijing, China*, J. A. Jacko and Marie-Claire Jenkins, Richard Churchill, Stephen Cox, and Dan Smith (eds.), Berlin: Springer, pp. 286-297.
- Graves, A., and Jaitly, N. (eds.) 2014. *Towards end-to-end speech recognition with recurrent neural networks*.
- Gregor, S. 2006. "The nature of theory in information systems," *MIS quarterly*, pp. 611-642.
- Gregor, S., and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS Quarterly* (37:2), pp. 337-355.
- Gregor, S., and Jones, D. 2007. "The anatomy of a design theory," *Journal of the Association for Information Systems* (8:5), p. 312.
- Gregor, S., Kruse, L. C., and Seidel, S. 2020. "The Anatomy of a Design Principle," *Journal of the Association for Information Systems*.
- Gribbons, B., and Herman, J. 1996. "True and quasi-experimental designs," *Practical assessment, research, and evaluation* (5:1), p. 14.

- Griol, D., Molina, J. M., and Callejas, Z. 2017. "Combining speech-based and linguistic classifiers to recognize emotion in user spoken utterances," *Neuro-computing*.
- Grudin, J. 1994. "Computer-supported cooperative work: History and focus," *Computer* (27:5), pp. 19-26.
- Guo, Y. R., and Goh, D. H.-L. 2015. "Affect in Embodied Pedagogical Agents: Meta-Analytic Review," *Journal of Educational Computing Research* (53:1), pp. 124-149.
- Gupta, S., and Bostrom, R. 2013. "Research Note—An Investigation of the Appropriation of Technology-Mediated Training Methods Incorporating Enactive and Collaborative Learning," *Information Systems Research* (24:2), pp. 454-469.
- Gupta, S., and Bostrom, R. P. 2009. "Technology-mediated learning: A comprehensive theoretical model," *Journal of the Association for Information Systems* (10:9), p. 686.
- Gupta, S., Bostrom, R. P., and Huber, M. 2010. "End-user training methods: What we know, need to know," *ACM SIGMIS Database* (41:4), pp. 9-39.
- Hair, J. F., Sarstedt, M., Hopkins, L., and Kuppelwieser, V. 2014. "Partial least squares structural equation modeling (PLS-SEM)," *European Business Review* (26:2), pp. 106-121.
- Halverson, L. R., and Graham, C. R. 2019. "Learner Engagement in Blended Learning Environments: A Conceptual Framework," *Online Learning* (23:2), pp. 145-178.
- Hannafin, M. J., Kim, M. C., and Kim, H. 2004. "Reconciling research, theory, and practice in web-based teaching and learning: The case for grounded design," *Journal of Computing in Higher Education* (15:2), p. 3.
- Harley, J. M., Bouchet, F., and Azevedo, R. 2012. "Measuring learners' co-occurring emotional responses during their interaction with a pedagogical agent in MetaTutor," in *International Conference on Intelligent Tutoring Systems*, Berlin, Heidelberg: Springer.
- Harsley, R., Di Eugenio, B., Green, N., Fossati, D., and Acharya, S. 2016. "Integrating Support for Collaboration in a Computer Science Intelligent Tutoring System," in *Intelligent Tutoring Systems: 13th International Conference, ITS 2016, Zagreb, Croatia, June 7-10, 2016. Proceedings*, A. Micarelli, J. Stamper and K. Panourgia (eds.), Cham, s.l.: Springer International Publishing, pp. 227-233.
- Hartson, R., and Pyla, P. S. 2012. *The UX Book: Process and guidelines for ensuring a quality user experience*, Elsevier.

- Hattie, J. 2012. *Visible learning for teachers: Maximizing impact on learning*, Routledge.
- Hauswald, J., Laurenzano, M. A., Zhang, Y., Li, C., Rovinski, A., Khurana, A., Dreslinski, R. G., Mudge, T., Petrucci, V., Tang, L., and Mars, J. 2015. "Sirius: An open end-to-end voice and vision personal assistant and its implications for future warehouse scale computers," in *ACM SIGPLAN Notices*, ACM SIGPLAN Notices (ed.), ACM, pp. 223-228.
- Hedegaard, M. 2002. "The zone of proximal development as basis for instruction," in *An introduction to Vygotsky*, Routledge, pp. 183-207.
- Heldner, M., and Edlund, J. 2010. "Pauses, gaps and overlaps in conversations," *Journal of Phonetics* (38:4), pp. 555-568.
- Hepburn, A., and Bolden, G. B. 2013. "The conversation analytic approach to transcription," *The handbook of conversation analysis* (57), p. 76.
- Hernández-Sellés, N., Muñoz-Carril, P.-C., and González-Sanmamed, M. 2019. "Computer-supported collaborative learning: An analysis of the relationship between interaction, emotional support and online collaborative tools," *Computers & Education* (138), pp. 1-12.
- Hesse, F., Care, E., Buder, J., Sassenberg, K., and Griffin, P. 2015. "A framework for teachable collaborative problem solving skills," in *Assessment and teaching of 21st century skills*, Springer, pp. 37-56.
- Hevner, A. R. 2007. "A three cycle view of design science research," *Scandinavian journal of information systems* (19:2), p. 4.
- Hevner, A. R., March, S. T., Park, J., and Ram, S. 2004. "Design science in information systems research," *MIS quarterly* (28:1), pp. 75-105.
- Hill, J., Randolph Ford, W., and Farreras, I. G. 2015. "Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations," *Computers in Human Behavior* (49), pp. 245-250.
- Hmelo-Silver, C. E. 2004. "Problem-based learning: What and how do students learn?" *Educational Psychology Review* (16:3), pp. 235-266.
- Hobert, S., and Wolff, R. M. von 2019. "Say Hello to Your New Automated Tutor—A Structured Literature Review on Pedagogical Conversational Agents," in *International Conference on Information Systems*, Munich, Germany.
- Hoey, E. M., and Kendrick, K. H. 2017. "Conversation analysis," *Research methods in psycholinguistics: A practical guide*, pp. 151-173.

- Hone, K. S., and El Said, G. R. 2016. "Exploring the factors affecting MOOC retention: A survey study," *Computers & Education* (98), pp. 157-168.
- Hooshyar, D., Ahmad, R. B., Yousefi, M., Fathi, M., Horng, S.-J., and Lim, H. 2016. "Applying an online game-based formative assessment in a flowchart-based intelligent tutoring system for improving problem-solving skills," *Computers & Education* (94), pp. 18-36.
- Hughes, C. 2017. 2018: *The Year of the Chatbot* – *Chatbots Magazine*. <https://chatbotsmagazine.com/2018-the-year-of-the-chatbot-fc5a5f780a31>. Accessed 8 January 2018.
- Hutchby, I., and Wooffitt, R. 2008. *Conversation analysis*, Polity.
- Ibrahim, M., and Al-Shara, O. (eds.) 2007. *Impact of interactive learning on knowledge retention*, Springer.
- ICEF 2019. *The year in MOOCs: Increase revenue and more degrees*. <http://monitor.icef.com/2019/01/the-year-in-moocs-increased-revenue-and-more-degrees/>. Accessed 13 October 2020.
- Im, I., and Hars, A. 2007. "Does a one-size recommendation system fit all? The effectiveness of collaborative filtering based recommendation systems across different domains and search modes," *ACM Transactions on Information Systems (TOIS)* (26:1), p. 4.
- Im Craik, F., and Lockhart, R. S. 1972. "Levels of processing: A framework for memory research," *Journal of verbal learning and verbal behavior* (11:6), pp. 671-684.
- iMotions 2019. *Facial Expression Analysis*. <https://imotions.com/biosensor/feafacial-expression-analysis/>. Accessed 2 December 2020.
- Janson, A., Söllner, M., and Leimeister, J. M. 2017. "Individual appropriation of learning management systems: Antecedents and consequences," *AIS Transactions on Human-Computer Interaction* (9:3), pp. 193-201.
- Jean-Charles, A. 2018. "Internet of Things in Education: Artificial Intelligence Voice Assistant in the Classroom," in *Proceedings of Society for Information Technology & Teacher Education International Conference*, Proceedings of Society for Information Technology & Teacher Education International Conference (ed.), Washington, D.C., United States, Association for the Advancement of Computing in Education (AACE).
- Jeong, H., and Hmelo-Silver, C. E. 2016. "Seven Affordances of Computer-Supported Collaborative Learning: How to Support Collaborative Learning? How Can Technologies Help?" *Educational Psychologist* (51:2), pp. 247-265 (doi: 10.1080/00461520.2016.1158654).

- Jonassen, D. 2011. "Supporting Problem Solving in PBL," *Interdisciplinary Journal of Problem-Based Learning* (5:2), pp. 95-119 (doi: 10.7771/1541-5015.1256).
- Jonassen, D., Davidson, M., Collins, M., Campbell, J., and Haag, B. B. 1995. "Constructivism and computer-mediated communication in distance education," *American journal of distance education* (9:2), pp. 7-26.
- Jonassen, D. H. 1991. "Evaluating constructivistic learning," *Educational Technology* (31:9), pp. 28-33.
- Jonassen, D. H. 2000. "Toward a design theory of problem solving," *Educational Technology Research and Development* (48:4), pp. 63-85.
- Kerly, A., Ellis, R., and Bull, S. 2008. "CALMsystem: A Conversational Agent for Learner Modelling," *Knowledge-Based Systems* (21:3), pp. 238-246.
- Kerry, A., Ellis, R., and Bull, S. 2009. "Conversational Agents in E-Learning," in *Applications and Innovations in Intelligent Systems XVI*, T. Allen, R. Ellis and M. Petridis (eds.), London: Springer London, pp. 169-182.
- Kim, B. 2001. "Social constructivism," *Emerging perspectives on learning, teaching, and technology* (1:1), p. 16.
- Kim, M. C., and Hannafin, M. J. 2011. "Scaffolding problem solving in technology-enhanced learning environments (TELEs): Bridging research and theory with practice," *Computers & Education* (56:2), pp. 403-417.
- Kim, R., Olfman, L., Ryan, T., and Eryilmaz, E. 2014. "Leveraging a personalized system to improve self-directed learning in online educational environments," *Computers & Education* (70), pp. 150-160.
- Kim, Y., and Baylor, A. L. 2016. "based design of pedagogical agent roles: A review, progress, and recommendations," *International Journal of Artificial Intelligence in Education* (26:1), pp. 160-169.
- King, A. 1993. "From sage on the stage to guide on the side," *College teaching* (41:1), pp. 30-35.
- Kirk, R. E. 1982. *Experimental design*, Wiley Online Library.
- Kizilcec, R. F., Papadopoulos, K., and Sritanyaratana, L. 2014. "Showing face in video instruction: effects on information retention, visual attention, and affect," in *Proceedings of the SIGCHI14 conference systems. ACM*.
- Klein, H. K., and Myers, M. D. 1999. "A set of principles for conducting and evaluating interpretive field studies in information systems," *MIS quarterly*, pp. 67-93.

- Klein, S. B., and Loftus, J. 1988. "The nature of self-referent encoding: The contributions of elaborative and organizational processes," *Journal of personality and Social Psychology* (55:1), p. 5.
- Kloos, C. D., Alario-Hoyos, C., Munoz-Merino, P. J., Ibanez, M. B., Estevez-Ayres, I., and Fernandez-Panadero, C. 2020. "Educational Technology in the Age of Natural Interfaces and Deep Learning," *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje* (15:1), pp. 26-33.
- Kock, W. D. de 2016. "Speech versus text supported hints in learning to solve word problems," *Computers in Human Behavior* (57), pp. 300-311.
- Kulik, J. A., and Fletcher, J. D. 2016. "Effectiveness of Intelligent Tutoring Systems," *Review of educational research* (86:1), pp. 42-78.
- Lajoie, S. P., and Azevedo, R. 2006. "Teaching and learning in technology-rich environments," *Routledge*.
- Larkin, J., Simon, H. 1987. "Why a Diagram is (Sometimes) Worth Ten Thousand Words," *Cognitive science* (11:1), pp. 65-100.
- Latham, A., Crockett, K., and McLean, D. 2014. "An adaptation algorithm for an intelligent natural language tutoring system," *Computers & Education* (71), pp. 97-110.
- Latham, A., Crockett, K., McLean, D., and Edmonds, B. 2012. "A conversational intelligent tutoring system to automatically predict learning styles," *Computers & Education* (59:1), pp. 95-109.
- Latham, A. M., Crockett, K. A., McLean, D. A., Edmonds, B., and O'shea, K. 2010. "Oscar: An intelligent conversational agent tutor to estimate learning styles," in *International Conference on Fuzzy Systems*, Barcelona, Spain. 2010, IEEE, pp. 1-8.
- Lederman, D. 2018. *Online Education Ascends*. <https://www.insidehighered.com/digital-learning/article/2018/11/07/new-data-online-enrollments-grow-and-share-overall-enrollment>.
- Leidner, D. E., and Jarvenpaa, S. L. 1995. "The use of information technology to enhance management school education: A theoretical view," *MIS quarterly*, pp. 265-291.
- Leimeister, J. M., Ebner, W., and Krcmar, H. 2005. "Design, implementation, and evaluation of trust-supporting components in virtual communities for patients," *Journal of Management Information Systems* (21:4), pp. 101-131.
- Lepadatu, I. 2012. "Use self-talking for learning progress," *Procedia - Social and Behavioral Sciences* (33), pp. 283-287.

- Lepp, M., Luik, P., Tauno, P., Kaspar, P., Reelika, S., Merilin, S., and Tõnisson, E. (eds.) 2017. *MOOC in programming: A success story*.
- Lepper, M. R., Drake, M. F., and O'Donnell-Johnson, T. "Scaffolding techniques of expert human tutors," in *Advances in learning & teaching. Scaffolding student learning: Instructional approaches and issues*, K. Hoga and M. Pressley (eds.), pp. 108-144.
- Levene, H. 1960. "Contributions to probability and statistics," *Essays in honor of Harold Hotelling*, pp. 278-292.
- Lewin, K. 1943. "Psychology and the process of group living," *The Journal of Social Psychology* (17:1), pp. 113-131.
- Limerick, H., Coyle, D., and Moore, J. W. 2014. "The experience of agency in human-computer interactions: a review," *Frontiers in human neuroscience* (8), p. 643.
- Lin, L., Atkinson, R. K., Christopherson, R. M., Joseph, S. S., and Harrison, C. J. 2013. "Animated agents and learning: Does the type of verbal feedback they provide matter?" *Computers & Education* (67), pp. 239-249.
- Lin, L., Ginns, P., Wang, T., and Zhang, P. 2020. "Using a pedagogical agent to deliver conversational style instruction: What benefits can you obtain?" *Computers & Education* (143), p. 103658.
- Lip ko, H. 2016. *Meet Jill Watson: Georgia Tech's first AI teaching assistant*. <https://pe.gatech.edu/blog/meet-jill-watson-georgia-techs-first-ai-teaching-assistant>. Accessed 5 January 2018.
- Litman, D. J., Rosé, C. P., Forbes-Riley, K., Vanlehn, K., Bhembé, D., and Silliman, S. 2004. "Spoken Versus Typed Human and Computer Dialogue Tutoring," in *Intelligent Tutoring Systems: 7th International Conference, ITS 2004, Maceió, Alagoas, Brazil, August 30 - September 3, 2004. Proceedings*, J. C. Lester, R. M. Vicari and F. Paraguaçu (eds.), Berlin, Heidelberg: Springer, pp. 368-379.
- Litman, D. J., Rosé, C. P., Forbes-Riley, K., Vanlehn, K., Bhembé, D., and Silliman, S. 2006. "Spoken versus typed human and computer dialogue tutoring," *International Journal of Artificial Intelligence in Education* (16:2), pp. 145-170.
- Liu, C.-W., Lowe, R., Serban, I. V., Noseworthy, M., Charlin, L., and Pineau, J. 2016. "How NOT to evaluate your dialogue system: An empirical study of unsupervised evaluation metrics for dialogue response generation," *arXiv preprint arXiv:1603.08023*.
- Llopart, A. 2020. "LiftFormer: 3D Human Pose Estimation using attention models," *arXiv preprint arXiv:2009.00348*.

- Lucas, G. M., Gratch, J., King, A., and Morency, L.-P. 2014. "It's only a computer: Virtual humans increase willingness to disclose," *Computers in Human Behavior* (37:5), pp. 94-100.
- Luckin, R., Holmes, W., Griffiths, M., and Forcier, L. B. 2016. "Intelligence unleashed: An argument for AI in education," *Pearson Education, London*.
- Lundqvist, K. O., Pursey, G., and Williams, S. (eds.) 2013. *Design and implementation of conversational agents for harvesting feedback in eLearning systems*, Springer.
- Lykken, D. T., and Venables, P. H. 1971. "Direct measurement of skin conductance: A proposal for standardization," *Psychophysiology* (8:5), pp. 656-672.
- Mackey, A., and Gass, S. M. 2011. *Research methods in second language acquisition: A practical guide*, John Wiley & Sons.
- Maedche, A., Morana, S., Schacht, S., Werth, D., and Krumeich, J. 2016. "Advanced user assistance systems," *Business & Information Systems Engineering* (58:5), pp. 367-370.
- Matsuda, N., Weng, W., and Wall, N. 2020. "The Effect of Metacognitive Scaffolding for Learning by Teaching a Teachable Agent," *International Journal of Artificial Intelligence in Education*, pp. 1-37.
- Mavrikis, M., Grawemeyer, B., Hansen, A., and Gutierrez-Santos, S. 2014. "Exploring the Potential of Speech Recognition to Support Problem Solving and Reflection," in *Open Learning and Teaching in Educational Communities*, C. Rensing, S. de Freitas, T. Ley and P. J. Muñoz-Merino (eds.), Cham: Springer International Publishing, pp. 263-276.
- Mayer, R. E., and Alexander, P. A. (eds.) 2011. *Handbook of research on learning and instruction*, New York [N.Y.]: Routledge.
- Maynard, D. W., and Clayman, S. E. 2003. "Ethnomethodology and conversation analysis," *Handbook of symbolic interactionism*, pp. 173-202.
- Mcalister, S., Ravenscroft, A., and Scanlon, E. 2004. "Combining interaction and context design to support collaborative argumentation using a tool for synchronous CMC," *Journal of Computer Assisted Learning* (20:3), pp. 194-204.
- McGrath, J. E. 1995. "Methodology matters: Doing research in the behavioral and social sciences," in *Readings in Human-Computer Interaction*, Elsevier, pp. 152-169.
- McKinsey & Company 2016. *Winning the expectations game in customer care*. <https://www.mckinsey.com/business-functions/operations/our-insights/winning-the-expectations-game-in-customer-care>. Accessed 1 May 2018.

- McNeal, K. S., Spry, J. M., Mitra, R., and Tipton, J. L. 2014. "Measuring student engagement, knowledge, and perceptions of climate change in an introductory environmental geology course," *Journal of Geoscience Education* (62:4), pp. 655-667.
- McTear, M., Callejas, Z., and Griol, D. 2016. "The conversational interface," *Springer* (6:94), p. 102.
- McTear, M. F. 2017. "The Rise of the Conversational Interface: A New Kid on the Block?" in *Future and Emerging Trends in Language Technology. Machine Learning and Big Data*, J. F. Quesada, F.-J. Martín Mateos and T. López Soto (eds.), Cham: Springer International Publishing, pp. 38-49.
- Meier, A., Spada, H., and Rummel, N. 2007. "A rating scheme for assessing the quality of computer-supported collaboration processes," *International Journal of Computer-Supported Collaborative Learning* (2:1), pp. 63-86.
- Merikivi, J., Salovaara, A., Mäntymäki, M., and Zhang, L. 2018. "On the way to understanding binge watching behavior: the over-estimated role of involvement," *Electronic Markets* (28:1), pp. 111-122.
- Merrill, M. D., Drake, L., Lacy, M. J., Pratt, J., and Group, I. R. 1996. "Reclaiming instructional design," *Educational Technology* (36:5), pp. 5-7.
- Mertens, W., Pugliese, A., and Recker, J. 2017. "Quantitative data analysis," *A companion*.
- Merton, R. K., and Merton, R. C. 1968. *Social theory and social structure*, Simon and Schuster.
- Miller, A. H., Feng, W., Fisch, A., Lu, J., Batra, D., Bordes, A., Parikh, D., and Weston, J. 2017. "Parlai: A dialog research software platform," *CoRR*.
- Mills, B., Evens, M., and Freedman, R. 2004. "Implementing directed lines of reasoning in an intelligent tutoring system using the atlas planning environment," in *International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004*, Las Vegas, NV, USA. 2004, IEEE, 729-733 Vol.1.
- Mills, K. L. (ed.) 2003. *Computer-supported cooperative work*, Citeseer.
- Molenaar, I., Roda, C., van Boxtel, C., and Sleegers, P. 2012. "Dynamic scaffolding of socially regulated learning in a computer-based learning environment," *Computers & Education* (59:2), pp. 515-523.
- Moreno, R. 2006. "Does the modality principle hold for different media? A test of the method-affects-learning hypothesis," *Journal of Computer Assisted Learning* (22:3), pp. 149-158.

- Mou, Y., and Xu, K. 2017. "The media inequality: Comparing the initial human-human and human-AI social interactions," *Computers in Human Behavior* (72), pp. 432-440.
- Mousavinasab, E., Zarifsanaiey, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., and Ghazi Saeedi, M. 2018. "Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods," *Interactive Learning Environments* (11:4), pp. 1-22.
- Mueller, B., and Urbach, N. 2017. "Understanding the why, what, and how of theories in IS research," *Communications of the Association for Information Systems* (41:1), p. 17.
- Murphy, N., and Messer, D. 2000. "Differential benefits from scaffolding and children working alone," *Educational Psychology* (20:1), pp. 17-31.
- Nass, C., Moon, Y., Morkes, J., Kim, E.-Y., and Fogg, B. J. 1997. "Computers are social actors: A review of current research," *Human values and the design of computer technology* (72), pp. 137-162.
- Nass, C., Steuer, J., and Tauber, E. R. 1994. "Computers are social actors," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, B. Adelson (ed.), Boston, Massachusetts, United States. 24.04.1994 - 28.04.1994, New York, NY: ACM, pp. 72-78.
- Nelson, L. M. 1999. "Collaborative problem solving," *Instructional design theories and models: A new paradigm of instructional theory* (2), pp. 241-267.
- Nicol, D. J., and Macfarlane-Dick, D. 2006. "Formative assessment and self-regulated learning: A model and seven principles of good feedback practice," *Studies in higher education* (31:2), pp. 199-218.
- Norman, M. A., and Thomas, P. J. 1991. "Informing HCI design through conversation analysis," *International journal of man-machine studies* (35:2), pp. 235-250.
- Novielli, N., Mazzotta, I., Carolis, B. D., and Pizzutilo, S. 2012. "Analysing user's reactions in advice-giving dialogues with a socially intelligent ECA," *Cognitive processing* (13 Suppl 2), pp. 487-497.
- Novielli, N., Rosis, F. de, and Mazzotta, I. 2010. "User attitude towards an embodied conversational agent: Effects of the interaction mode," *Journal of Pragmatics* (42:9), pp. 2385-2397.
- Nye, B. D. 2014. "Barriers to ITS Adoption: A Systematic Mapping Study," in *Intelligent Tutoring Systems*, D. Hutchison, T. Kanade, J. Kittler, J. M. Kleinberg, A. Kobsa, F. Mattern, J. C. Mitchell, M. Naor, O. Nierstrasz, C. Pandu Rangan, B. Steffen, D. Terzopoulos, D. Tygar, G. Weikum, S. Trausan-Matu, K. E.

- Boyer, M. Crosby and K. Panourgia (eds.), Cham: Springer International Publishing, pp. 583-590.
- Nye, B. D., Graesser, A. C., and Hu, X. 2014. "AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring," *International Journal of Artificial Intelligence in Education* (24:4), pp. 427-469.
- O'Brien, L. 1989. "Learning styles: Make the student aware," *NASSP Bulletin* (73:519), pp. 85-89.
- Oeste, S., Lehmann, K., Janson, A., Söllner, M., and Leimeister, J. M. 2015. "Re-designing University Large Scale Lectures: How To Activate The Learner," in *75th Academy of Management Annual Meeting*, Vancouver, British Columbia.
- Orlikowski, W. J., and Baroudi, J. J. 1991. "Studying information technology in organizations: Research approaches and assumptions," *Information Systems Research* (2:1), pp. 1-28.
- Ormerod, R. J. 2009. "The history and ideas of critical rationalism: the philosophy of Karl Popper and its implications for OR," *Journal of the Operational Research Society* (60:4), pp. 441-460.
- O'shea, K. 2014. "Natural language scripting within conversational agent design," *Applied Intelligence* (40:1), pp. 189-197.
- Oudeyer, P.-Y., Gottlieb, J., and Lopes, M. 2016. "Chapter 11 - Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies," in *Progress in Brain Research : Motivation Theory, Neurobiology and Applications*, B. S. a. S. Knecht (ed.), Elsevier, pp. 257-284.
- Oxley, N. L., Dzindolet, M. T., and Paulus, P. B. 1996. "The effects of facilitators on the performance of brainstorming groups," *Journal of social behavior and personality* (11:4), pp. 633-646.
- Pai, K.-C., Kuo, B.-C., Liao, C.-H., and Liu, Y.-M. 2020. "An application of Chinese dialogue-based intelligent tutoring system in remedial instruction for mathematics learning," *Educational Psychology* (35:2), pp. 1-16.
- Parviainen, E., and Søndergaard, M. L. J. 2020. "Experiential Qualities of Whispering with Voice Assistants," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, R. Bernhaupt, F. 'F.' Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjørn, S. (S.) Zhao, B. P. Samson and R. Kocielnik (eds.), Honolulu HI USA. 25 04 2020 30 04 2020, New York, NY, USA: ACM, pp. 1-13.
- Pearl, C. 2016. *Designing Voice User Interfaces: Principles of Conversational Experiences*, " O'Reilly Media, Inc."

- Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A design science research methodology for information systems research," *Journal of Management Information Systems* (24:3), pp. 45-77.
- Pelk, H. 2016. *Machine learning, neural networks and algorithms – ITNEXT*. <https://itnext.io/machine-learning-neural-networks-and-algorithms-4cd27a105b89>. Accessed 6 January 2018.
- Peña, E. A., and Slate, E. H. 2006. "Global validation of linear model assumptions," *Journal of the American Statistical Association* (101:473), pp. 341-354.
- Pereira, J. 2016. "Leveraging chatbots to improve self-guided learning through conversational quizzes," in *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality*, Salamanca, Spain: ACM, pp. 911-918.
- Perez, S. 2017. *Voice-enabled smart speakers to reach 55% of U.S. households by 2022*. <https://techcrunch.com/2017/11/08/voice-enabled-smart-speakers-to-reach-55-of-u-s-households-by-2022-says-report/>. Accessed 12 November 2019.
- Pérez, J., Cerezo, E., Serón, F. J., and Rodríguez, L.-F. 2016. "A cognitive-affective architecture for ECAs," *Biologically Inspired Cognitive Architectures* (18), pp. 33-40.
- Pickard, M. D., Roster, C. A., and Chen, Y. 2016. "Revealing sensitive information in personal interviews: Is self-disclosure easier with humans or avatars and under what conditions?" *Computers in Human Behavior* (65), pp. 23-30.
- PISA 2017. *How does PISA measure students' ability to collaborate?*
- Popenici, S. A. D., and Kerr, S. 2017. "Exploring the impact of artificial intelligence on teaching and learning in higher education," *Research and Practice in Technology Enhanced Learning* (12:1), p. 22.
- Popper, K. R. 1959. "The propensity interpretation of probability," *The British journal for the philosophy of science* (10:37), pp. 25-42.
- Porcheron, M., Fischer, J. E., McGregor, M., Brown, B., Luger, E., Candello, H., and O'Hara, K. 2017. "Talking with Conversational Agents in Collaborative Action," in *Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing - CSCW '17 Companion*, C. P. Lee, S. Poltrock, L. Barkhuus, M. Borges and W. Kellogg (eds.), Portland, Oregon, USA. 2/25/2017 - 3/1/2017, New York, New York, USA: ACM Press, pp. 431-436.
- Porcheron, M., Fischer, J. E., Reeves, S., and Sharples, S. 2018. "Voice Interfaces in Everyday Life," in *Proceedings of the 2018 CHI Conference on Human*

Factors in Computing Systems - CHI '18, Montreal QC, Canada. 4/21/2018 - 4/26/2018, New York, New York, USA: ACM Press, pp. 1-12.

- Recker, J. 2013. *Scientific research in information systems: A beginner's guide*, Berlin, Heidelberg: Springer.
- Reeves, B., and Nass, C. I. 1996. *The media equation: How people treat computers, television, and new media like real people and places*, Cambridge University Press.
- Richardson, J., and Swan, K. 2003. "Examining social presence in online courses in relation to students' perceived learning and satisfaction," *Faculty and Staff Publications and Research - Center for Online Learning, Research and Service (COLRS)* (1:3).
- Rietsche, R., Duss, K., Persch, J. M., and Soellner, M. 2018. "Design and Evaluation of an IT-based Formative Feedback Tool to Foster Student Performance," in *39th International Conference on Information Systems (ICIS)*, San Francisco, USA, San Francisco, CA, USA, pp. 1-17.
- Romiszowski, A. J., and Haas, J. A. de 1989. "Computer mediated communication for instruction: Using e-mail as a seminar," *Educational Technology* (29:10), pp. 7-14.
- Roschelle, J. 2013. "Special issue on CSCL: discussion," *Educational Psychologist* (48:1), pp. 67-70.
- Roschelle, J., and Teasley, S. D. 1995. "The Construction of Shared Knowledge in Collaborative Problem Solving," in *Computer Supported Collaborative Learning*, C. O'Malley (ed.), Berlin, Heidelberg: Springer, pp. 69-97.
- Rosé, C., and Vanlehn, K. 2005. "An evaluation of a hybrid language understanding approach for robust selection of tutoring goals," *International Journal of Artificial Intelligence in Education* (15:4), pp. 325-355.
- Ruan, S., Jian, L., Xu, J., Joe-Kun Tham, B., Qiu, Z., Zhu, Y., Murnane, E. L., Brunskill, E., and Landay, J. A. 2019. "QuizBot: A Dialogue-based Adaptive Learning System for Factual Knowledge," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (ed.), ACM.
- Rummel, N., and Spada, H. 2005. "Learning to collaborate: An instructional approach to promoting collaborative problem solving in computer-mediated settings," *Journal of the Learning Sciences* (14:2), pp. 201-241.
- Runes, D. D. 1984. *Dictionary of philosophy: Revised and enlarged*, AbeyBooks.
- Ryan, G. W., and Bernard, H. R. 2003. "Techniques to identify themes," *Field methods* (15:1), pp. 85-109.

- Saadatzai, M. N., Pennington, R. C., Welch, K. C., Graham, J. H., and Scott, R. E. 2017. "The Use of an Autonomous Pedagogical Agent and Automatic Speech Recognition for Teaching Sight Words to Students with Autism Spectrum Disorder," *Journal of Special Education Technology* (32:3), pp. 173-183.
- Sacks, H., Schegloff, E. A., and Jefferson, G. 1978. "A simplest systematics for the organization of turn taking for conversation," in *Studies in the organization of conversational interaction*, Elsevier, pp. 7-55.
- Sahin, D., and Yilmaz, R. M. 2020. "The effect of Augmented Reality Technology on middle school students' achievements and attitudes towards science education," *Computers & Education* (144), p. 103710 (doi: 10.1016/j.compedu.2019.103710).
- Sarker, S., Xiao, X., and Beaulieu, T. 2013. "Guest editorial: qualitative studies in information systems: a critical review and some guiding principles," *MIS quarterly* (37:4), pp. iii-xviii.
- Schegloff, E. A. 1997. "Practices and actions: Boundary cases of other-initiated repair," *Discourse processes* (23:3), pp. 499-545.
- Schunk, D. H. 1991. "Self-efficacy and academic motivation," *Educational Psychologist* (26:3-4), pp. 207-231.
- Seaman, J. E., Allen, I. E., and Seaman, J. 2018. *Grade Increase: Tracking Distance Education in the United States*. <https://www.onlinelearningsurvey.com/highered.html>. Accessed 14 March 2020.
- Seeber, I., Bittner, E., Briggs, R. O., Vreede, G.-J. de, and Vreede, T. 2018. "Machines as Teammates: A Collaboration Research Agenda," in *Hawaii International Conference on System Sciences (HICSS)*, Hawaii, USA.
- Senn, S. 1994. "Testing for baseline balance in clinical trials," *Statistics in medicine* (13:17), pp. 1715-1726.
- Serban, I. V., Sankar, C., Germain, M., Zhang, S., Lin, Z., Subramanian, S., Kim, T., Pieper, M., Chandar, S., and Ke, N. R. 2017. *A deep reinforcement learning chatbot*, Montreal.
- Setlock, L. D., Fussell, S. R., and Neuwirth, C. 2004. "Taking it out of context," in *Proceedings of the 2004 ACM conference on Computer supported cooperative work - CSCW '04*, J. Herbsleb and G. Olson (eds.), Chicago, Illinois, USA. 06.11.2004 - 10.11.2004, New York, New York, USA: ACM Press, p. 604.
- Sherhoff, D. J., Csikszentmihalyi, M., Schneider, B., and Sherhoff, E. S. 2014. "Student engagement in high school classrooms from the perspective of flow theory," in *Applications of flow in human development and education*, Springer, pp. 475-494.

- Siadaty, M., Gašević, D., and Hatala, M. 2016. "Measuring the impact of technological scaffolding interventions on micro-level processes of self-regulated workplace learning," *Computers in Human Behavior* (59), pp. 469-482.
- Sinatra, G. M., Heddy, B. C., and Lombardi, D. 2015. *The challenges of defining and measuring student engagement in science*, Taylor & Francis: Taylor & Francis.
- Situated design for creative, reflective, collaborative, technology-mediated learning* 2014.
- Soderstrom, N. C., and Bjork, R. A. 2015. "Learning versus performance: An integrative review," *Perspectives on Psychological Science* (10:2), pp. 176-199.
- Söllner, M., Bitzer, P., Janson, A., and Leimeister, J. M. 2017. "Process is king: Evaluating the performance of technology-mediated learning in vocational software training," *Journal of Information Technology* (18:2), p. 159.
- Song, D., Oh, E. Y., and Rice, M. 2017. "Interacting with a conversational agent system for educational purposes in online courses," in *2017 10th International Conference on Human System Interactions (HSI): Proceedings : International Hall, University of Ulsan, Ulsan, Republic of Korea, July 17-19, 2017*, Ulsan, South Korea. 7/17/2017 - 7/19/2017, Piscataway, NJ: IEEE, pp. 78-82.
- Song, Y., Wong, L.-H., and Looi, C.-K. 2012. "Fostering personalized learning in science inquiry supported by mobile technologies," *Educational Technology Research and Development* (60:4), pp. 679-701.
- Sonnenberg, C., and Vom Brocke, J. (eds.) 2012. *Evaluations in the science of the artificial—reconsidering the build-evaluate pattern in design science research*, Springer.
- Sonwalkar, N. 2008. "Adaptive individualization: The next generation of online education," *On the horizon* (16:1), pp. 44-47.
- Sottolare, R. A., Graesser, A. C., Hu, X., Olney, A., Nye, B., and Sinatra, A. M. 2016. *Design Recommendations for Intelligent Tutoring Systems: Volume 4-Domain Modeling*, US Army Research Laboratory.
- Spiro, R. J., Feltovich, P. J., Feltovich, P. L., Jacobson, M. J., and Coulson, R. L. 1991. "Cognitive flexibility, constructivism, and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains," *Educational Technology* (31:5), pp. 24-33.
- Spradley, J. P. 2016. *The ethnographic interview*, Waveland Press.
- Tam, M. 2000. "Constructivism, instructional design, and technology: Implications for transforming distance learning," *Educational Technology & Society* (3:2), pp. 50-60.

- Tarouco, L. M. R., Silveira, C., and Krassmann, A. L. 2018. "Collaborative Learning with Virtual Entities," in *Learning and Collaboration Technologies. Design, Development and Technological Innovation: 5th International Conference, LCT 2018, Held as Part of HCI International 2018, Las Vegas, NV, USA, July 15-20, 2018, Proceedings, Part I*, P. Zaphiris and A. Ioannou (eds.), Cham: Springer International Publishing, pp. 480-493.
- Taylor, J. 2006. "Evaluating mobile learning: What are appropriate methods for evaluating learning in mobile environments," *Big issues in mobile learning* (1:3), pp. 25-27.
- Team, R. C. 2013. *R: A language and environment for statistical computing*. <http://www.R-project.org/>. Accessed 1 May 2019.
- Tegos, S., and Demetriadis, S. 2017. "Conversational Agents Improve Peer Learning through Building on Prior Knowledge," *Educational Technology & Society* (20:1), pp. 99-111.
- Torrance, H. 2012. "Triangulation, respondent validation, and democratic participation in mixed methods research," *Journal of mixed methods research* (6:2), pp. 111-123.
- Tukey, J. W. 1949. "Comparing individual means in the analysis of variance," *Biometrics* (5:2), pp. 99-114.
- Urquhart, C., Lehmann, H., and Myers, M. D. 2010. "Putting the 'theory' back into grounded theory: guidelines for grounded theory studies in information systems," *Information Systems Journal* (20:4), pp. 357-381.
- Vail, A. K., Wiggins, J. B., Grafsgaard, J. F., Boyer, K. E., Wiebe, E. N., and Lester, J. C. 2016. "The Affective Impact of Tutor Questions: Predicting Frustration and Engagement," *International Educational Data Mining Society* (2:9), pp. 247-254.
- van Aken, J. E. 2004. "Management research based on the paradigm of the design sciences: the quest for field-tested and grounded technological rules," *Journal of management studies* (41:2), pp. 219-246.
- van der Meij, H., van der Meij, J., and Harmsen, R. 2015. "Animated pedagogical agents effects on enhancing student motivation and learning in a science inquiry learning environment," *Educational Technology Research and Development* (63:3), pp. 381-403.
- van Raaij, E. M., and Schepers, J. J. L. 2008. "The acceptance and use of a virtual learning environment in China," *Computers & Education* (50:3), pp. 838-852.

- Vanlehn, K. 2011. "The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems," *Educational Psychologist* (46:4), pp. 197-221.
- Venable, J., and Baskerville, R. 2012. "Eating our own cooking: Toward a more rigorous design science of research methods," *Electronic Journal of Business Research Methods* (10:2).
- Venable, J., Pries-Heje, J., and Baskerville, R. 2016. "FEDS: a framework for evaluation in design science research," *European Journal of Information Systems* (25:1), pp. 77-89.
- Venkatesh, V., and Bala, H. 2008. "Technology acceptance model 3 and a research agenda on interventions," *Decision Sciences* (39:2), pp. 273-315.
- Venkatesh, V., Brown, S. A., and Bala, H. 2013. "Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems," *MIS quarterly*, pp. 21-54.
- Venkatesh, V., Brown, S. A., and Sullivan, Y. W. 2016. "Guidelines for conducting mixed-methods research: An extension and illustration," *Journal of the Association for Information Systems* (17:7), p. 2.
- Vincent, and James 2016. *Twitter taught Microsoft's AI chatbot to be a carist asshole in less than a day*. <https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist>. Accessed 6 January 2018.
- VoiceFlow 2019. *Design, prototype and build voice apps*. <https://www.voiceflow.com/>. Accessed 1 December 2020.
- Vom Brocke, J., and Liang, T.-P. 2014. "Guidelines for neuroscience studies in information systems research," *Journal of Management Information Systems* (30:4), pp. 211-234.
- Vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., and Cleven, A. 2015. "Standing on the Shoulders of Giants: Challenges and Recommendations of Literature Search in Information Systems Research," *CAIS* (37), p. 9.
- Vygotsky, L. S. 1978. *Mind in society: The development of higher mental process*, Cambridge, MA: Harvard University Press.
- W3C Community and Business Groups 2019. *Speech API*. <https://w3c.github.io/speech-api/>. Accessed 11 October 2020.
- Wallace, R. S. 2007. *The Anatomy of ALICE*, University Press: University Press.

- Wambsganss, T., Niklaus, C., Cetto, M., Söllner, M., Handschuh, S., and Leimeister, J. M. 2020. "AL: An Adaptive Learning Support System for Argumentation Skills," in *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, R. Bernhaupt, F. 'F.' Mueller, D. Verweij, J. Andres, J. McGrenere, A. Cockburn, I. Avellino, A. Goguey, P. Bjørn, S. (S.) Zhao, B. P. Samson and R. Kocielnik (eds.), Honolulu HI USA. 25 04 2020 30 04 2020, New York, NY, USA: ACM, pp. 1-14.
- Wang, D., Han, H., Zhan, Z., Xu, J., Liu, Q., and Ren, G. 2015. "A problem solving oriented intelligent tutoring system to improve students' acquisition of basic computer skills," *Computers & Education* (81), pp. 102-112.
- Weber, R. 2012. "Evaluating and developing theories in the information systems discipline," *Journal of the Association for Information Systems* (13:1), p. 2.
- Webster, J., and Watson, R. T. 2002. "Analyzing the past to prepare for the future: Writing a literature review," *MIS quarterly*, pp. 13-23.
- Weizenbaum, J. 1966. "ELIZA—a computer program for the study of natural language communication between man and machine," *Communications of the ACM* (9:1), pp. 36-45.
- Wells, J. D., Campbell, D. E., Valacich, J. S., and Featherman, M. 2010. "The effect of perceived novelty on the adoption of information technology innovations: A risk/reward perspective," *Decision Sciences* (41:4), pp. 813-843.
- Wieling, M. B., and Hofman, W. H.A. 2010. "The impact of online video lecture recordings and automated feedback on student performance," *Computers & Education* (54:4), pp. 992-998.
- Wik, P., and Hjalmarsson, A. 2009. "Embodied conversational agents in computer assisted language learning," *Speech Communication* (51:10), pp. 1024-1037.
- Winkler, R., Hobert, S., and Appius, M. 2020a. "Empowering Educators to Build their Own Conversational Agents in Online Learning," in *Proceedings of the 2020 European Conference of Information Systems*, Marrakesh, Morocco.
- Winkler, R., Hobert, S., Salovaara, A., Söllner, M., and Leimeister, J. M. 2020b. "Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent," in *ACM CHI Conference on Human Factors in Computing Systems (CHI)*, Honolulu, Hawaii.
- Winkler, R., Neuweiler, M. L., Bittner, E., and Söllner, M. 2019a. "Hey Alexa, Please Help Us Solve This Problem! How Interactions with Smart Personal Assistants Improve Group Performance," in *Proceedings of the International Conference on Information Science*, Munich, Germany.

- Winkler, R., and Roos, J. 2019. "Bringing AI into the Classroom: Designing Smart Personal Assistants as Learning Tutors," in *International Conference on Information Systems*, Munich, Germany.
- Winkler, R., and Söllner, M. "Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis," in *Academy of Management Annual Meeting*, Chicago, IL, USA.
- Winkler, R., and Söllner, M. 2020. "Towards Empowering Educators to Create their own Smart Personal Assistants," in *Hawaii International Conference on System 2020*, Maui, HI, USA.
- Winkler, R., Söllner, M., and Leimeister, J. M. 2019b. "Improving Students' Problem-Solving Skills with Smart Personal Assistants," in *International Conference on Information Systems*, Munich, Germany.
- Winkler, R., Weingart, P., and Söllner, M. 2019c. "Using Smart Personal Assistants for Online Learning Activities: What benefits can we expect?" in *Wirtschaftsinformatik*, Siegen, Germany.
- Witte, E. H. 2007. "Toward a group facilitation technique for project teams," *Group Processes & Intergroup Relations* (10:3), pp. 299-309.
- Wood, D., Bruner, J. S., and Ross, G. 1976. "The role of tutoring in problem solving," *Journal of child psychology and psychiatry* (17:2), pp. 89-100.
- Woolf, B. P., Aïmeur, E., Nkambou, R., and Lajoie, S. (eds.) 2008. *Intelligent Tutoring Systems*, Berlin, Heidelberg: Springer Berlin Heidelberg.
- Woolfolk Hoy, A. 2000. "Educational psychology in teacher education," *Educational Psychologist* (35:4), pp. 257-270.
- Xie, H., Chu, H.-C., Hwang, G.-J., and Wang, C.-C. 2019. "Trends and development in technology-enhanced adaptive/personalized learning: A systematic review of journal publications from 2007 to 2017," *Computers & Education* (140), p. 103599.
- Xolocotzin Eligio, U., Ainsworth, S. E., and Crook, C. K. 2012. "Emotion understanding and performance during computer-supported collaboration," *Computers in Human Behavior* (28:6), pp. 2046-2054.
- Xu, J. D., Benbasat, I., and Cenfetelli, R. T. 2014. "The Nature and Consequences of Trade-off Transparency in the Context of Recommendation Agents," *MIS quarterly* (38:2).
- Yang, J. X., Unhelkar, V. V., Li, K., and Shah, J. A. 2017. "Evaluating effects of user experience and system transparency on trust in automation," in *Evaluating effects of user experience and system transparency on trust in automation*, Chicago, USA: IEEE.

- Yvonne Feilzer, M. 2010. "Doing mixed methods research pragmatically: Implications for the rediscovery of pragmatism as a research paradigm," *Journal of mixed methods research* (4:1), pp. 6-16.
- Zawacki-Richter, O., Marín, V. I., Bond, M., and Gouverneur, F. 2019. "Systematic review of research on artificial intelligence applications in higher education – where are the educators?" *International Journal of Educational Technology in Higher Education* (16:1), p. 7228.

Appendix

Appendix A Appendices to the study presented in Section 5.2

Appendix A.1 Lessons and learning goals

Lesson (15 lessons in total per class)	Learning Goals
Lesson 1	LG1: Students should understand the necessity of law in everyday life.
Lessons 2 to 4	LG2: Students should explain differences between morality, custom, and law.
Lessons 5 to 7	LG3: Students should solve problems related to the freedom of opinion.
Lessons 8 to 10	LG4: Students should solve problems related to the freedom of religion.
Lessons 11 to 13	LG5: Students should solve problems related to the property guarantee.
Lessons 14 to 15	LG6: Students should analyze in which cases fundamental right have their limitations.

Appendix A.2 Homework assignment 2 and characteristics of a problem-based task

Homework Assignment 2	
Task: Imagine the following environment: The use of smartphones during the lesson is forbidden in your school. Nevertheless, Thomas K. (a classmate of yours) uses his smartphone during the lesson to text his mother that he will be late that day. The teacher confiscates the smartphone and tells Thomas K. that she will keep the smartphone until the end of the lesson. How would you solve this emerging problem with the help of the law?	
Characteristics of a problem-task (Jonassen 2000)	How I addressed the characteristics
Ill-structured	The task can be considered as ill-structured, because not every piece of information is given. Moreover, it gives no advice on how to solve the task.

Open-ended	The task leaves room for interpretation. There is no single right answer. The students have to interpret the legal articles in a correct way.
Realistic	The task relates to a topic that is currently highly discussed in the relevant country. It can therefore be considered as realistic.
Resonate with the executors' experience	Since every participant owns a smartphone and has experienced similar situations in their own class, they can put themselves in Thomas K.'s position.

Appendix A.3 Pre-survey items

Pre-survey		
Variable	Item	Scale
Pre-experience with SPAT	1. Have you ever used a Smart Personal Assistant (e.g. Amazon's Alexa, Google's Assistant, Apple's Siri)? 2. If yes, how often do you use a Smart Personal Assistant per lesson on average?	Yes/No open
Personal innovativeness	1. If I heard about a new information technology, I would look for ways to experiment with it. 2. Among my peers, I am usually the first to try out new information technologies.	1 to 7 (7 = highest) 1 to 7 1 to 7

	3. In general, I am hesitant in trying out new information technologies (reverse-scored). 4. I like to experiment with new information technologies.	1 to 7
Demographics	1. Age 2. Gender 3. Apprenticeship company (for vocational business school only)	open open open

Appendix A.4 Detailed pre-test results and student characteristics

	Experiment Class (SPA)			Control Class (traditional learning materials)		
Study 1: High School						
	Points achieved	Std. Dev.	N	Mean	Std. Dev.	N
Sub-task 1 (total: 12)	2.65	1.32	23	2.60	1.94	22
Sub-task 2 (total: 12)	3.31	1.59	23	3.21	2.53	22
Sub-task 3 (total: 12)	4.50	2.11	23	4.46	2.83	22
Total	10.46	1.67	23	10.27	2.44	23
Study 2: Vocational Business School						
	Points achieved	Std. Dev.	N	Mean	Std. Dev.	N
Sub-task 1 (total: 12)	2.85	1.25	23	2.55	1.54	22
Sub-task 2 (total: 12)	4.00	1.65	23	3.55	2.34	22
Sub-task 3 (total: 12)	4.85	2.02	23	4.60	2.12	22

Total	11.70	1.63	23	10.67	2.16	23
--------------	--------------	-------------	-----------	--------------	-------------	-----------

Appendix A.5 Pretest and posttest tasks

Learning goal	Pretest	Posttest
LG3: Solve problems related to the freedom of opinion	Imagine that in your country, political parties issue posters that are obviously against foreigners. Some foreigners complain about this. How would you solve this problem?	Imagine that in your country, Michael P, a good friend of yours, issues flyers that are obviously against one of the left-wing politicians in the country. Some people complain about this. How would you solve this problem?
LG4: Solve problems related to the freedom of religion	Imagine that in your country, women are not allowed to wear burqas in public. Adem and Merve, two Islamic women, do not care about this rule and go out with their burqas. They are caught by the police and have to pay a fee now. They are complaining about it. How would you solve this problem?	Imagine that Muslims in your country have built a mosque in your neighborhood. After a while, a neighbor gets excited and criticizes the construction of the mosque. How would you solve the problem?
LG5: Solve problems related to property guarantee	Imagine the state you are living in wants to build a road on your land. If you do not allow this, your land will be expropriated. You are filling a complaint against the expropriation of your land. How would you solve this problem?	Imagine that the city you are living in want to create a one-week food festival and, thus, needs two of your land plots, against which you are filing a complaint. How would you solve this problem?

Appendix A.6 General rating framework per task

Problem-solving step	Key Questions (1 Point per question)	Points
1. Problem identification and engagement	What is the main problem involved here? Who are the involved parties?	2
2. Problem exploration	Which area of the law is affected? What are the interests of the individual parties?	2
3. Problem reconstruction	What is a possible solution for the problem? Which law article could apply here? What does the given law article say?	3
4. Solution presentation and communication	How can the law article be applied in the context? Are other articles also applicable? Who can raise which claims?	3
5. Reflection and Negotiation	How can you justify the solution? How can the related fundamental rights be restricted?	2
Total		12

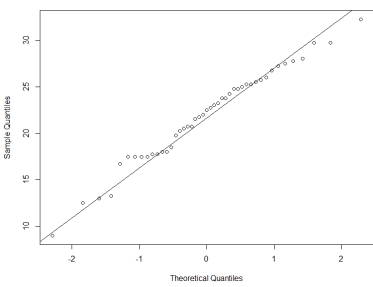
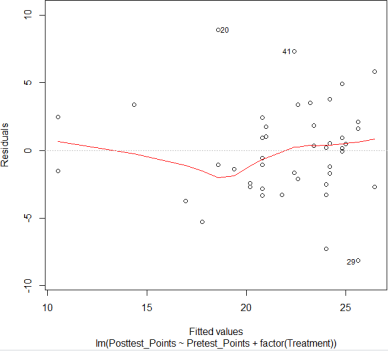
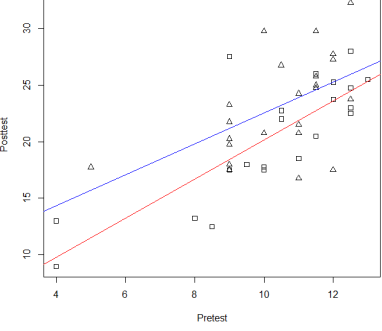
Appendix A.7 Post-survey items

Post-survey		
Variable	Item	Scale
Question 1	To what extent do you feel that the learning materials have helped you? Explain your answer	open

Question 2 (treatment-specific)	What are your experiences when using the Smart Personal Assistant as a tutor?	open
------------------------------------	---	------

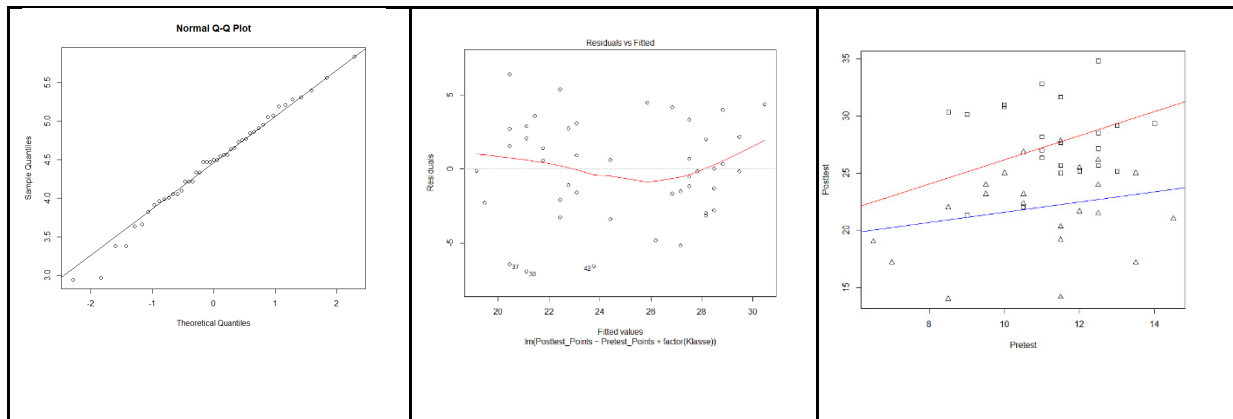
Appendix A.8 Check for assumptions for ANCOVA

Experiment 1 in high school

Statistical and visual test for normality	Test for homogeneity of variance	Test for homogeneity of regression slopes
Shapiro-Wilk normality test $W = 0.98714$ $p\text{-value} = 0.8927$	Levene's test $F = 0.0861$ $p\text{-value} = 0.7706$	Interaction analysis $F = 0.277$ $p\text{-value} = 0.601$
		

Experiment 2 in vocational business school

Tests for normality	Tests for homogeneity of variance	Tests for homogeneity of regression slopes
Shapiro-Wilk normality test $W = 0.98408$ $p\text{-value} = 0.7847$	Levene's test $F = 0.1299$ $p\text{-value} = 0.7203$	Interaction analysis $F = 1.305$ $p\text{-value} = 0.260$



Appendix B Appendices to the study presented in Section 5.4

Appendix B.1 Introductory text and corresponding questions about SWOT analysis

“SWOT is a strategic planning technique used to help a person or organization identify strengths, weaknesses, opportunities, and threats related to business competition or project planning.

The aim of SWOT is to specify the objectives of the business venture or project and identify the internal and external factors that are favorable and unfavorable to achieving those objectives.

SWOT consists of internal factors and external factors. The internal factors are the strengths and weaknesses internal to the organization. The external factors are the opportunities and threats presented by the environment external to the organization. Examples of internal factors are financial performance, human capital, know-how, image, customer loyalty, patents, corporate culture, leadership, human capital satisfaction, and innovativeness. Examples of external factors are competitors as well as technological, cultural, social, and political trends and changes. The difference between internal factors and external factors is that a business can actively influence internal factors, while a business for the most part can only react to the external factors that affect the company's environment. To apply a SWOT analysis, examples for each category for a given business need to be specified.”

Questions

- What is the difference between a strength and an opportunity?
- Please name a Strength for Amazon.
- Please name a Weakness for Volkswagen (the German diesel car manufacturer).
- Please name an Opportunity for SpaceX (the space company belonging to Elon Musk).
- Please name a Threat for Facebook.

Appendix C Appendices to the study presented in Section 6.2

Appendix C.1 Route code snippet for creating a SPAT

```
1  const { check } = require("express-validator");
2  const authMiddleware = require("../middleware/authMiddleware");
3
4  module.exports = (app) => {
5    const chatbots = require("../controllers/chatbot.controller.js");
6
7    // Retrieve all chatbots from current user
8    app.get("/api/chatbots", authMiddleware, chatbots.findAllFromUser);
9
10   // Retrieve all public chatbots from all users
11   app.get("/api/chatbots/public", authMiddleware, chatbots.getAllPublic);
12
13   // Create a new chatbot
14   app.post(
15     "/api/chatbots",
16     [
17       check("name", "Valid name is required.").isLength({ min: 5 }),
18       check("type", "Valid type is required.").isLength({ min: 5 }),
19       check("public", "Valid public is required.").exists(),
20       authMiddleware,
21     ],
22     chatbots.create
23   );
```

Appendix C.2 Controller code snippet for creating a SPAT

```
1  const Chatbot = require("../models/chatbot.model.js");
2  const Node = require("../models/node.model.js");
3  const History = require("../models/history.model.js");
4  const { validationResult } = require("express-validator");
5  const appconfig = require("../config/app.config.js");
6  const crypto = require("crypto");
7  const randomInt = require("random-int");
8
9  // Create a new chatbot
10 exports.create = (req, res) => {
11   // Validate request
12   const errors = validationResult(req);
13
14   if (!errors.isEmpty()) {
15     console.log(errors.array());
16     return res.status(422).jsonp({
17       message: errors.array()[0].msg,
18     });
19   }
20 }
```

Appendix C.3 Model code snippet for creating a SPAT

```
1  const sql = require("./db.js");
2
3  // constructor
4  const Chatbot = function(chatbot) {
5    this.name = chatbot.name;
6    this.type = chatbot.type;
7    this.public = chatbot.public;
8    this.link = chatbot.link;
9    this.owner = chatbot.owner;
10   this.start_node = chatbot.start_node;
11   this.pin = chatbot.pin;
12 };
13
14 Chatbot.create = (newchatbot, result) => {
15   sql.query("INSERT INTO chatbots SET ?", newchatbot, (err, res) => {
16     if (err) {
17       console.log("error: ", err);
18       result(err, null);
19       return;
20     }
21
22     console.log("created chatbot: ", { id: res.insertId, ...newchatbot });
23     result(null, { id: res.insertId, ...newchatbot });
24   });
25 }
```

Appendix C.4 Constructs for summative evaluation of SPAT platform

Constructs	Questionnaire Construct
Personal innovativeness	1. When I hear about a new information technology, I look for ways to try it. 2. Among my colleagues I am usually the first to test new information technologies. 3. I enjoy experimenting with new information technologies. 4. In general, I am reluctant to try out new information technologies.
Interactional enjoyment	1. It is fun and enjoyable to have a conversation with my specially created chatbot. 2. I was very engrossed in the conversation with the chatbot. 3. The conversation with the specially created chatbot is exciting. 4. I could imagine that learners often prefer to communicate with the chatbot than with the teacher. 5. The services offered by the chat-

	bot are more entertaining and attractive than without a chatbot.
User satisfaction	1. I was satisfied with the individual steps needed to create a chatbot as provided on the website. 2. I am satisfied with the service offered by the website. 3. The interaction with the website was a pleasant and satisfying experience. 4. The individual steps included on the website gave me useful information. 5. The answers from the specially created chatbot were suitable / appropriate. 6. Overall, I would describe the conversation with the specially created chatbot as satisfactory.
Intention to use	1. I will use the website again to create chatbots. 2. I would recommend the website for creating chat bots to others. 3. If this website for creating chatbots was available in stores, I would buy it. 4. I would like the website for creating chatbots to help me make decisions. 5. I am satisfied with the use of the website for creating chatbots because it is easy to use and better than preparing traditional learning materials. 6. If I had the chance to use this chatbot again, I would do so.
Perceived usefulness	1. Using the website to create chatbots in my work would allow me to complete tasks faster. 2. Using the website to create chatbots would improve my performance. 3. Using the website to create chatbots in my work would improve my productivity.

	4. Using the website to create chatbots would increase the effectiveness of my job. 5. Using the website to create chatbots would make my job easier. 6. I would find the chatbot creation site useful in my job.
Ease of use	1. Learning how to use the website to create chatbots is easy for me. 2. I find it easy to get the chatbot creation site to do what I want. 3. My interaction with the chatbot creation site was clear and easy to understand. 4. I find this chatbot creation site clear and easy to understand. 5. It is easy for me to use this product skillfully. 6. I find this chatbot creation site easy to use.
User experience test	To rate the platform, please give your opinion on the following pairs of contrasting characteristics: unpleasant – pleasant; incomprehensible – understandable; creative – unimaginative; easy to learn – difficult to learn; valuable – useless; boring – exciting; uninteresting – interesting; intuitive – unpredictable; fast – slow; innovative – conventional; constructive – hindering; good – bad; complicated – easy; safe – unsafe; activating – soporific; etc.

CURRICULUM VITAE

Personal details and contact information

Name Rainer Winkler
Date of Birth 06/29/1990
Nationality Austria
Address Dammstrasse 17, 9230 Flawil
Tel.: +41 71 224 32 15
Mob. Tel.: +41 76 336 36 76
Mail: rainer.winkler@unisg.ch



Education and Qualifications

since 08/2017 PHD candidate at University of St. Gallen
Title of Dissertation: Understanding and Designing Smart Personal Assistants in Online and Blended Learning Environments
Supervisor: Prof. Dr. Jan Marco Leimeister

10/2015 – 07/2017 University of St. Gallen
Additional Degree in Business Education

10/2013 – 06/2015 University of Innsbruck
Master in Business Education

10/2010 – 07/2013 University of Innsbruck
Bachelor in Business Education

10/2006 – 07/2010 Handelsakademie Landeck (AT)
Specialization in IT

Work experience

Since 01/2017 University of St. Gallen
Research Assistant and Member of the Program management in Business Administration

07/2015 – 07/2017 Sportgymnasium Davos
Berufsschule Wattwil
Teacher in Business and Law

IT-Skills Java, JavaScript, Python, R
React and MaterialUI
Microsoft Office