

**HTW** Chur

Hochschule für Technik und Wirtschaft  
University of Applied Sciences

## Churer Schriften zur Informationswissenschaft

Herausgegeben von  
Wolfgang Semar

---

Arbeitsbereich  
Informationswissenschaft

**Schrift 99**

# Enhancing Collaboration in Collaborative Problem-Solving with Conversational Agents

Hanna Kummel

---

Chur 2019

**Churer Schriften zur Informationswissenschaft**

Herausgegeben von Wolfgang Semar

Schrift 99

## Enhancing Collaboration in Collaborative Problem-Solving with Conversational Agents

Hanna Kummel

Diese Publikation entstand im Rahmen einer Thesis zum Master of Science FHO  
in Business Administration, Major Information and Data Management.

Referent: Prof. Dr. Ingo Barkow

Korreferent: Dr. Heiko Rölke

**Verlag:** Arbeitsbereich Informationswissenschaft

**ISSN:** 1660-945X

**Chur,** März 2019

## Abstract

OECD has contributed a rich and solid framework to the field of collaborative problem-solving. The framework was developed due to the integration of a collaborative problem-solving assessment in PISA 2015. A recognizable restriction to the field is the exclusion of natural language which prevents implementation of a more realistic scenario. The maintenance of the validity of the item today forces the assessment setup to retain multiple-choice questions. This circumstance has been widely criticized and thus became the focus of this master's thesis. Subsequently, the main research question to answer is:

*What influence does the implementation of a "natural collaboration" have on the assessments for collaborative problem-solving?*

The seven chapters of this thesis will answer this question with conducted empirical research and a synthesis of results. Results led to a functional prototype of a conversational agent that resolves the linguistic limitation for further research.

**Keywords:** collaborative problem-solving, conversational agents, natural language processing

## Kurzfassung

Die OECD hat einen umfangreichen und soliden Rahmen für die kollaborative Problemlösung geschaffen. Das Framework wurde aufgrund der Integration in die Erhebung kollaborativen Problemlösens in PISA 2015 entwickelt. Eine erkennbare Einschränkung des Feldes ist der Ausschluss der natürlichen Sprache, was die Umsetzung als realistischeres Szenario verhindert. Die Erhaltung der Validität des Messinstrumentes zwingt den Versuchsaufbau dazu Multiple-Choice-Fragen zu nutzen. Dieser Umstand wurde vielfach kritisiert und stellt damit den Fokus dieser Masterarbeit dar. In der Folge lautet die Hauptforschungsfrage:

*Welchen Einfluss hat die Umsetzung einer "natürlichen Zusammenarbeit" auf die Erhebung kollaborativer Problemlösungen?*

Die sieben Kapitel dieser Arbeit werden diese Frage durch empirische Forschung und einer Synthese der Ergebnisse beantworten. Die Ergebnisse führen zu einem funktionalen Prototyp eines Konversationsagenten, der die sprachliche Einschränkung für weitere Forschungen auflöst.

**Schlagworte:** kollaboratives Problemlösen, Konversationsagenten, natürliche Sprachverarbeitung

## Index

Abstract.....	i
Kurzfassung .....	i
List of Figures.....	iv
List of Tables.....	v
Abbreviations .....	vi
1 Introduction .....	1
1.1 Purpose .....	2
1.2 Outline .....	2
1.3 Research Methodology .....	3
2 Literature Review .....	5
2.1 Collaborative Problem-Solving.....	6
2.2 Conversational Agents .....	13
2.3 Synthesis of the Literature Analysis .....	15
3 Empirical Analysis: Expert Interviews.....	17
3.1 Methodological Approach.....	17
3.2 Operationalization .....	17
3.2.1 Sampling.....	18
3.2.2 Experts .....	18
3.2.3 Data Processing and Analysis.....	18
3.3 Analysis of the Expert Interview .....	19
3.3.1 Description of Results .....	19
3.3.2 Discussion of the Results .....	28
3.4 Interview Summary .....	32
4 Assumptions from Empirical Research.....	33
4.1 Referencing the Research Questions.....	33
4.2 Criteria for the Development of an Agent .....	34
5 Prototyping the Agent.....	37

---

5.1	Process.....	37
5.2	Execution .....	38
5.2.1	Idea-Generation Stakeholder Requirements .....	38
5.2.2	Matching System Requirements .....	42
5.2.3	Development System Architectural Design .....	43
5.2.4	Implementation .....	44
5.3	Description of System Integration.....	48
5.4	Reflection of the Prototype's Creation .....	49
6	Results of this Master's Thesis.....	51
7	Discussion and Outlook .....	53
8	References .....	55
9	Appendix A – Interview Data.....	59
10	Appendix B – Technical Documentation.....	111

## List of Figures

Figure 1: DRM Framework by Blessing and Chakrabarti .....	3
Figure 2 Research Design of this master's thesis following DRM .....	4
Figure 3: P21 Framework for 21st Century Learning .....	7
Figure 4: Conceptual framework for collaborative problem-solving .....	10
Figure 5: OECD Xandar Item, "Part 1, Item 4: Agreeing on Strategy" .....	11
Figure 6: Chatbot Architecture of a conversational AI chatbot .....	15
Figure 7: Criterion 1 - Implementation of natural language .....	35
Figure 8: Criterion 2 - Implementation of interactivity .....	35
Figure 9: Criterion 3 - Implementation of strong characteristics .....	36
Figure 10: ISO/IEC 12207 "Systems and software engineering — Software life cycle processes" .....	38
Figure 11: Chat Client Telegram.....	39
Figure 12: Interface of Xandar Unit in PISA 2015 marking the two main components of the interface: Conversation- and Problem Space .....	40
Figure 13: Extended Problem-Space in the Xandar Unit of PISA 2015 assessment.....	41
Figure 14: Mockup of the expected final state of the system .....	42
Figure 15: System components .....	44
Figure 16: Architecture of the interactive component.....	44
Figure 17: Intents in the Google Dialogflow system .....	47
Figure 18: Training Interface of the Google Dialogflow System .....	48
Figure 19: Xandar Unit simulation within a chat-based environment.....	49

## List of Tables

Table 1: Cooper's taxonomy applied to literature review process. ....	5
Table 2: Main features of the PISA problem-solving framework.....	8
Table 3: Matrix of collaborative problem-solving skills for PISA 2015 .....	11
Table 4: Persona 1 - Anna.....	114
Table 5: Persona 2 - Kevin .....	115

## Abbreviations

ABS	Agent Based System
ACTNext	Organization of American College Testing
ColPS	Collaborative Problem-solving
CSV	Comma Separated Values
JSON	JavaScript Object Notation
NLP	Natural Language Processing
NLU	Natural Language Understanding
OECD	Organisation for Economic Co-operation and Development
PIAAC	Programme for the International Assessment of Adult Competencies
PISA	Programme for International Student Assessment

## 1 Introduction

Collaborative problem-solving has been identified as a major 21st century skill by many significant institutions such as the OECD, which defined collaborative problem-solving competency (subsequently named ColPS) in 2013 and assessed it in its large-scale PISA studies in 2015:

*“Collaborative problem-solving competency is the capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution and pooling their knowledge, skills and efforts to reach that solution.”*  
(OECD 2013, p.6)

In 2009, 94 % of 921 industries in North America and Europe utilized or planned to utilize some form of web-based technology, including e-mail, videoconferencing, instant messaging and others, to facilitate collaborative problem-solving (Enterprise and SMB Software Survey, North America and Europe, Forrester report, 2009). This trend has a high significance to researchers from different fields such as educational research, psychology and computer science, as well as many sub-areas of these fields. Despite the fact that ColPS seems to offer an endless amount of discoverable phenomena, my research always leads back to the one restriction that concerned me most: the distance between current computer-based assessments and state-of-the-art “real world” technology. Discussion about this particular gap can be found in many publications of recent years; for example, Greiff, Wüstenberg, Holt, Goldhammer, and Funke (2013) wrote about the lack of assessment tools to assess complex problem-solving. By applying complex problem-solving items based on the microDYN framework into a current state-of-the-art assessment platform such as “CBA ItemBuilder”, Care & Griffin (2014) found that *“platform availability for test implementation, delivery, and scoring in combination with a well-founded theoretical concept enables educationally motivated research”*. Thus, the need for adequate software in order to succeed in enhancing educationally motivated research is clear. However, available software often does not fulfill all the requirements for the purposes for ColPS assessments. The development of computer-based assessment requires diverse skills such as educational and pedagogical skills, programming, user interface design and conception, server administration, and data management, in addition to research expertise and content knowledge (Rölke, 2012). Access to skills that are demanding a high-level of expertise is only one factor of complexity when it comes to the development of appropriate assessments of collaborative problem-solving.

Agent-based approaches, and especially conversational agents, are a rich field in technical-educational research. Many existing projects follow an agent-based approach including, for

example, AutoTutor, which was developed by the Institute for Intelligent Systems at the University of Memphis by A. Gresser (Person et al., 2007). Even though the PISA 2015 ColPS framework highlights the implementation of conversational agents (OECD, 2017), these have not yet been implemented in PISA 2015 ColPS assessments authentically.

This master's thesis is an applied research approach following the design research methodology (Blessing & Chakrabarti, 2009) to enrich the research field by proposing an agile conversational agent.

## **1.1 Purpose**

The purpose of this thesis is two-fold. The first aim is to evaluate the construct of ColPS assessment settings as human-human and human-agent systems. The second aim is to propose an approach that unifies advantages of both approaches of interaction (human-human and human-agent). The overall goal is to maintain the content validity of human-agent assessment by providing a prototype of an agent that allows participants to solve problems by using natural language. In the field of artificial intelligence (AI) and in the field of interpretation and structure of natural language processing (NLP) many technologies exist that could be of value. Given the title of this thesis, "Enhancing Collaboration in Collaborative Problem-solving with Conversational Agents", this thesis initially sets out to answer the following questions:

- What is the role of agent-based systems for collaborative problem-solving?
- What are the most common approaches and associated difficulties of assessing collaborative problem-solving?
- What influence does the implementation of a "natural collaboration" have on the assessments for collaborative problem-solving?

## **1.2 Outline**

This paper consists of seven chapters and is structured as follows. In the next chapter, the background of collaborative problem-solving is described and a review of earlier studies and their results given. Also, the theory behind conversational agents is examined more closely. To gain an extensive insight into the current state of the field, expert interviews were conducted with researchers in collaborative problem-solving and agent-based systems. These initial examinations were carried out to gain a deeper understanding of the field and its further development, as described in Chapter 3. Chapter 4 concludes the theoretical and qualitative examination and explains the interdependence of the constructs of ColPS and conversational agents. Based on these findings, research questions provide the information needed to define the scope of the prototyped environment. Established frameworks and items are taken as the foundation of the prototype, to give a clear direction to create a

comparable, valid scenario of CoIPS. Chapter 5 describes the components and the development of the prototype, a conversational agent environment that aims to serve as an experimental platform in order to unleash the CoIPS construct.

### 1.3 Research Methodology

The present work follows a design research methodology, which is a rather new research method suitable not only for actual design tasks but also for the development of scientific artefacts. This method has been proven to be a good approach for developing actual artefacts, especially for research in information technology. According to Blessing and Chakrabarti (2009, preface viii p.7) the essential ambition of the design research methodology (subsequently named DRM) is to help engineering and industrial design research to gain in relevance, efficiency and effectiveness. The operationalization will be executed as three superordinate steps, as recommended by the developers of the DRM framework (Figure 1).

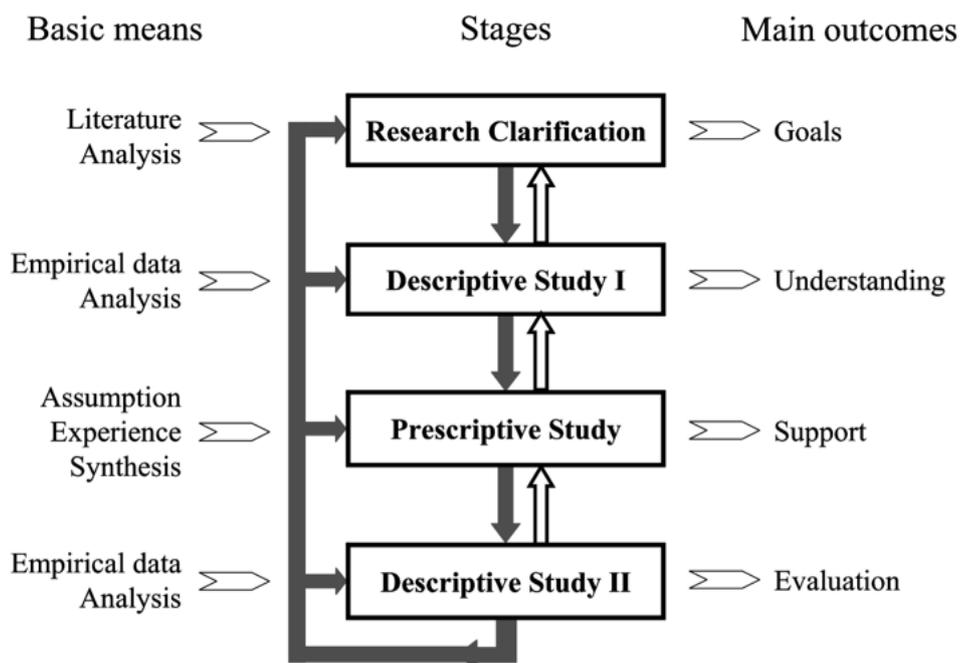


Figure 1: DRM Framework by Blessing and Chakrabarti (2009, p. 15)

The applied design research methodology is adjusted to the context of the thesis, as suggested in the DRM handbook (Blessing & Chakrabarti, 2009, p.17), and subsequently subdivided into six separate subtasks. These subtasks are: discover, define, conceptualize, design, implement and evaluate the implemented artefact. The graphic below (Figure 2) shows these steps, including relevant sub-steps, that need to be taken in order to achieve a viable prototype.

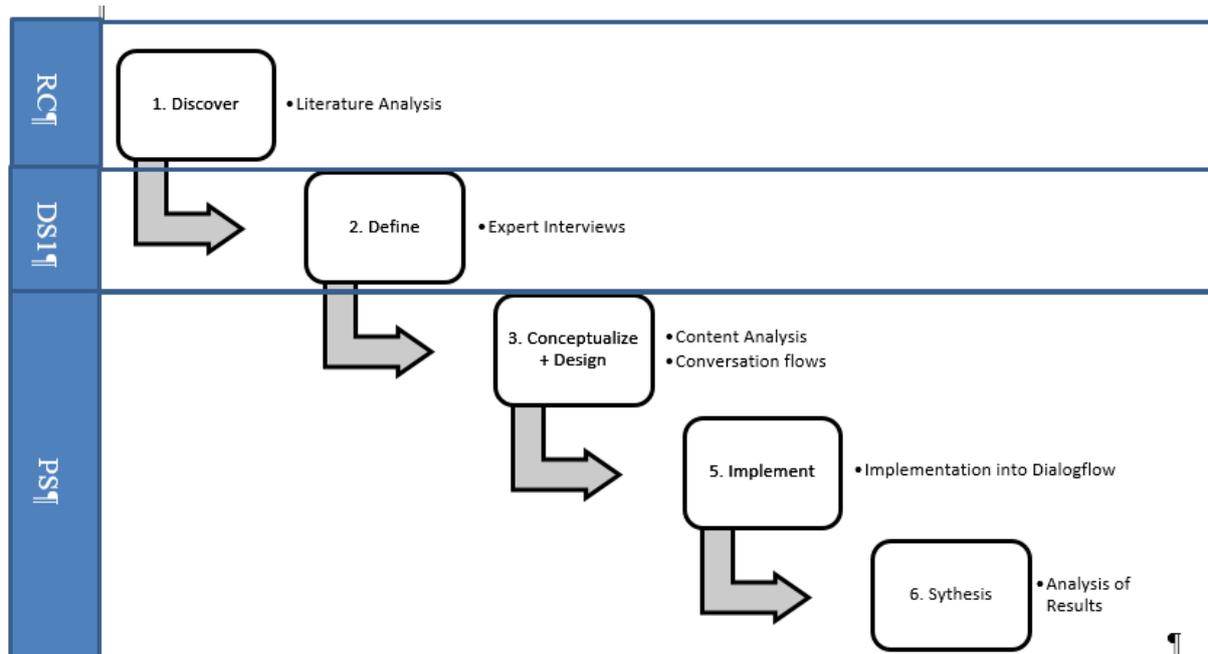


Figure 2 Research Design of this master's thesis following DRM (figure by author)

The scope that defines the research goal of this thesis is following the three suggested steps of the design research method, applied to a realistic scope for a master's thesis. The thesis begins with the research clarification (RC) that is carried out through a focused literature analysis following the suggested approach of vom Brocke et al. (2009). This stage determines the aim, focus and scope of the research project.

A comprehensive descriptive study (DS1) is the next logical step in order to identify the criteria of success for the prototype. This descriptive study is presented as an expert interview following the approach suggested by Bogner and Menz (2009, p.46) and was conducted with five experts of the field. Its analysis follows Mayring's (2000) findings for qualitative content analysis.

Subsequently, the identified criteria result in the conceptualization and development (PS) of the conversational agent. Blessing and Chakrabarti (2009) point out that publications falsely tend to end with a prescriptive study, whereas they should be followed by another descriptive study to evaluate the results (p.18). However, this approach applies to more extensive research papers such as dissertations (p.19).

## 2 Literature Review

This chapter presents the literature analysis, firstly by examining the collaborative problem-solving construct itself and its development in recent years and, secondly, by examining agent-based systems and, more specifically, conversational agents. In order to develop a new artefact for the field by opening the language space to the assessment construct, detailed descriptions of both fields are required. According to vom Brocke, Simons, Niehaves, Plattfaut, Cleven and Reimer (2009), it is advisable to structure a literature review following Cooper’s taxonomy (Cooper, 1988, cited in vom Brocke et al., 2009). This approach helps to filter and arrange the literature needed to answer the research question. Table 1 shows the focus of the literature review, which forms the basis of the creation of the interview questions in Chapter 3.

<i>Characteristics</i>	<i>Categories</i>			
<i>Focus</i>	Research Outcomes	Methods	Theories	Application
<i>Goal</i>	Integration	Criticism	Central issues	
<i>Organization</i>	Historical	Conceptual	Methodological	
<i>Perspective</i>	Neutral Representation		Espousal of Position	
<i>Audience</i>	Specialized scholars	General Scholars	Practitioners	General public
<i>Coverage</i>	Exhaustive	Exhaustive and selective	Representative	Central/Pivotal

Table 1: Cooper's taxonomy applied to literature review process.

Highlighted fields display the emphasized literature characteristics (own presentation)

A circumstance that is limiting the extent of the literature review is the novelty of both research fields, especially regarding the application-focused approach of this thesis. This circumstance sets the focus of the literature review as being mainly on methodological publications, as well as case studies and research reports concerning the application of ColPS as an assessment or as an assessment-independent construct. In line with the research goal of this thesis, publications that carry some sort of criticism or analysis of central issues are especially promising as they can give direction to the setting of the scope for development of the prototype. Most literature is of a conceptual nature, but methodological publications are also relevant to this review. Due to the high level of complexity, the audience of most publications comprises specialized scholars. However, the publications of the PISA 2015 ColPS Assessment by OECD is also aimed at general scholars. As mentioned above, the relevant literature is limited but the available literature has been examined exhaustively.

## 2.1 Collaborative Problem-Solving

A short introduction to the definition of collaborative problem-solving was given in the beginning of this thesis (Chapter 1). To understand better the evolution of ColPS, it is necessary to describe its relevance to society as a so-called “21<sup>st</sup>-century skill” and its general backdrop: problem-solving or individual problem-solving. This section analyzes the discussion of ColPS as a 21<sup>st</sup> century skill and the theory behind problem-solving; it then uses the results to inform the theory of collaborative problem-solving.

- **21<sup>st</sup> Century Skills**

It is obvious that people need to adjust to the increasing complexity in work and private lives brought about by the development in industries, technology and the digital sector. With the evolution of technology, economies have also shifted from industrial to information-based and knowledge-based (Griffin et al., 2012, p.2). This development led to the necessity of adjustment, not least in education. Initial thoughts about this shift and so-called 21<sup>st</sup> century skills can be traced back to the 1980s but came concrete with the initiation of ATC21s (Assessment and Teaching of 21st century) project; a 21<sup>st</sup> century skill can be any skill that is essential for navigating the 21<sup>st</sup> century. “*The project explores changing forms of assessment to match the conceptualization of twenty-first-century skills. It introduces a methodology for large scale innovative and technology-rich approaches to assessment.*” (Griffin, 2012, p.7) Skills that are seen as necessary and feasible include collaborative problem-solving and learning through a digital network. Besides ATC21s, there are other institutions focusing on a definition of required 21<sup>st</sup> century skills. One of them is P21, which proposes a wider view on the complex circumstances and defines skills as 4Cs: communication, collaboration, critical thinking and creativity (p21.org, last access 12.07.2018).

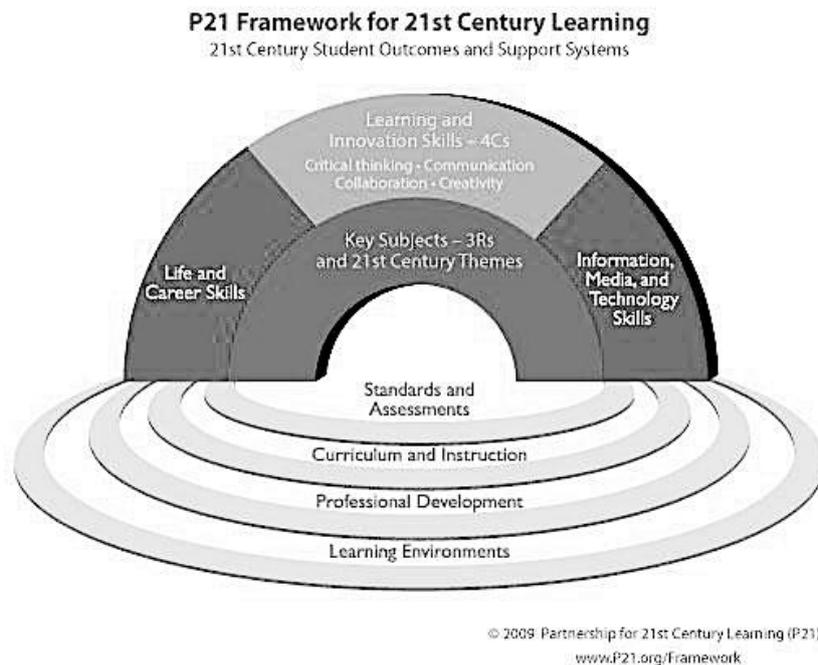


Figure 3: P21 Framework for 21st Century Learning (Partnership for 21st Century Skills, 2009)

P21's proposed framework (Figure 3) is subdivided into key subjects that they refer to as the 3Rs (reading, writing, arithmetic), life and career skills, learning and innovation skills (the 4Cs) and information, media and technology skills (Partnership for 21<sup>st</sup> Century Skills, 2009). While there are other frameworks that are not being mentioned here, the idea of 21<sup>st</sup> century skills is explained in order to underline the need for teaching communication and collaboration in education. Consequently, a further evolution of problem-solving and collaborative problem-solving skills is a great enhancement into forming a successful society.

#### • Problem-solving

Wang and Chiew describe a problem as a construct based on three elements: givens, goals and operations (cited in Ormrod, 1999; Polya, 1954). "Givens" are the available information that belongs to the problem, "goals" are the desired termination of the problem-state and "operations" are potential actions that can be executed to achieve the goals. Wang and Chiew (2008, p.3) also make clear the possibility that an individual might not find a solution to the problem, because many factors influence the individual's ability to solve a problem. A representative approach to a problem-solving procedure has been proposed by Polya (1954) and follows four steps:

- (A) Understanding the problem, which means to identify the "givens" of a problem.
- (B) Devising a plan that determines appropriate actions to solve the problem.
- (C) Carrying out the plan to execute the actions that have been determined in step (B).
- (D) Looking backward on the overall effectiveness of the approach to the problem.

At the end of this process, there should be a method for learning from the experience and recording the solution in order to repeat actions on similar problems in the future. This theoretical derivation of the problem-solving construct can be recognized within OECDs problem-solving framework which is explained below.

Before the ColPS assessment, OECD had already assessed problem-solving in the PISA studies in the years 2009 and 2012. ColPS is a further refinement of the problem-solving construct, which is why it is necessary to deduce from the theory of problem-solving first. PISA 2012 defines problem-solving as:

*...an individual's capacity to engage in cognitive processing to understand and resolve problem situations where a method of solution is not immediately obvious. It includes the willingness to engage with such situations in order to achieve one's potential as a constructive and reflective citizen. (OECD, 2014, p.30)*

The process of individual problem-solving follows four steps:

- (A) Gathering information related to the problem.
- (B) Representing the problem and the various relationships in the problem with tables, graphs, symbols or words.
- (C) Devising a strategy to solve the problem and carrying out this strategy.
- (D) Ensuring that the strategy has been followed and reacting to feedback obtained during the course of solving the problem (OECD, 2017, p.2).

These four steps are still relevant to the problem-solving aspects of the PISA 2015 collaborative problem-solving assessment described in the next paragraph.

<b>NATURE OF THE PROBLEM SITUATION</b> Is all the information needed to solve the problem disclosed at the outset?	▪ <i>Interactive</i> : not all information is disclosed; some information has to be uncovered by exploring the problem situation.		
	▪ <i>Static</i> : all relevant information for solving the problem is disclosed at the outset.		
<b>PROBLEM-SOLVING PROCESS</b> What are the main cognitive processes involved in the particular task?	▪ <i>Exploring and understanding</i> the information provided with the problem.		
	▪ <i>Representing and formulating</i> : constructing graphical, tabular, symbolic or verbal representations of the problem situation and formulating hypotheses about the relevant factors and relationships between them.		
	▪ <i>Planning and executing</i> : devising a plan by setting goals and sub-goals, and executing the sequential steps identified in the plan.		
	▪ <i>Monitoring and reflecting</i> : monitoring progress, reacting to feedback, and reflecting on the solution, the information provided with the problem, or the strategy adopted.		
<b>PROBLEM CONTEXT</b> In what everyday scenario is the problem embedded?	▪ <i>Setting</i> : does the scenario involve a technological device? <table border="1" style="margin-left: 20px;"> <tbody> <tr> <td>– <i>Technology</i> (involves a technological device)</td> </tr> <tr> <td>– <i>Non-technology</i></td> </tr> </tbody> </table>	– <i>Technology</i> (involves a technological device)	– <i>Non-technology</i>
	– <i>Technology</i> (involves a technological device)		
	– <i>Non-technology</i>		
	▪ <i>Focus</i> : what environment does the problem relate to? <table border="1" style="margin-left: 20px;"> <tbody> <tr> <td>– <i>Personal</i> (the student, family or close peers)</td> </tr> <tr> <td>– <i>Social</i> (the community or society in general)</td> </tr> </tbody> </table>	– <i>Personal</i> (the student, family or close peers)	– <i>Social</i> (the community or society in general)
– <i>Personal</i> (the student, family or close peers)			
– <i>Social</i> (the community or society in general)			

Table 2: Main features of the PISA problem-solving framework (OECD, 2012, p.31)

Table 2 shows the main features of the PISA problem-solving framework, which has been the foundation of the assessment since 2012; the main parts of this framework are the nature of the problem-solving situation, the problem-solving process itself, and the problem context. The two different representations of the problem-solving processes that are presented by PISA and P21 show a slight difference but are mostly following the same process.

- **Collaborative Problem-solving**

In the introduction of this thesis, CoIPS was defined, and it is the capacity of an individual to engage effectively in a process with two or more participants in order to solve a problem. For an effective problem-solving strategy based on collaboration, the group process is an elementary part.

As described in a reflexive article about the PISA CoIPS assessment 2015 (OECD, 2017), the framework of problem-solving that was defined for PISA 2012 has been resurrected and complemented with three further major skills that one may master specifically to solve problems collaboratively:

- (A) Establishing and maintaining shared understanding (finding out what other team members know and ensuring that team members share the same vision of the problem).
- (B) Taking appropriate action to solve the problem (determining what collaborative actions need to be performed).
- (C) Establishing and maintaining team organization (following one's own role in the problem-solving strategy and checking that others also follow their assigned role).

The construct of CoIPS has been generalized with several contexts in mind. One view on the topic is the separation of skills needed to solve a problem collaboratively (Figure 4). Griffin et al. divided the construct into two major skills, namely social skills and cognitive skills (2010). Within the social skills, a person has to understand how to participate, how to establish their own perspective and how to regulate actions socially. The ability to regulate tasks and build knowledge then leads to the cognitive skills that are required.

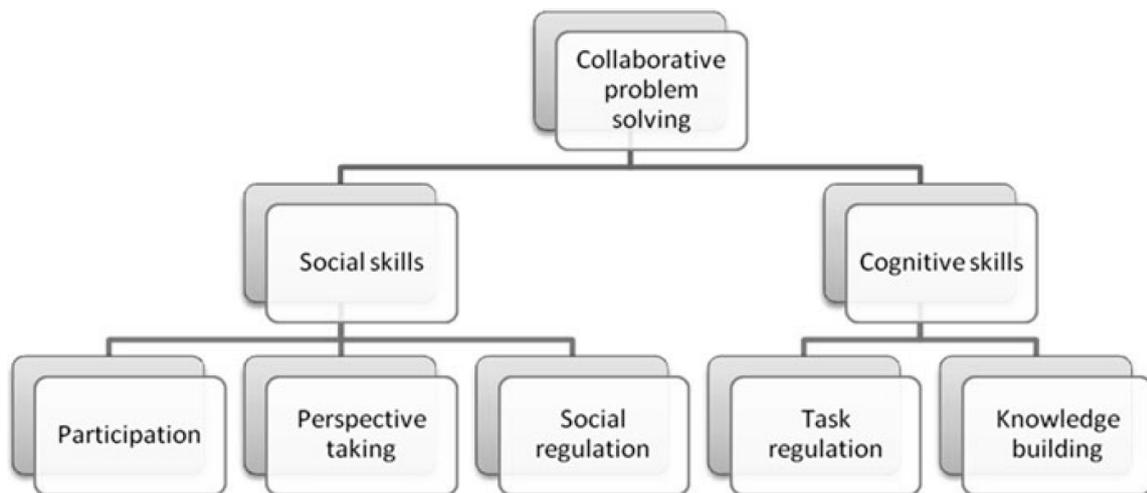


Figure 4: Conceptual framework for collaborative problem-solving (Source: Griffin et al. 2010)

- **Assessing collaborative problem-solving**

To date, CoIPS in PISA 2015 has been evaluated using computer-based assessments with static and pre-defined dialogue structures which are similar to multiple-choice questions. In this setup, the student interacts with a simulation that is quite unnatural. This fragmented state of development can be explained by the high complexity in all the different fields such as interpretation of natural language, the creation of extensive elaborated artificial characters and the preservation of the validity of the assessment itself.

OECD is considering CoIPS as a computer-based assessment with an agent-based system using conversational agents (OECD 2017, p.21ff). The CoIPS skillset (Table 3) has been tested and conducted as a computer-based assessment, but the component of the conversational agent has not been developed sufficiently; it uses a rather static construct with pre-defined dialogue structures which are similar to multiple-choice assessments.

	(1) Establishing and maintaining shared understanding	(2) Taking appropriate action to solve the problem	(3) Establishing and maintaining team organisation
(A) Exploring and understanding	(A1) Discovering perspectives and abilities of team members	(A2) Discovering the type of collaborative interaction to solve the problem, along with goals	(A3) Understanding roles to solve the problem
(B) Representing and formulating	(B1) Building a shared representation and negotiating the meaning of the problem (common ground)	(B2) Identifying and describing tasks to be completed	(B3) Describing roles and team organisation (communication protocol/rules of engagement)
(C) Planning and executing	(C1) Communicating with team members about the actions to be/being performed	(C2) Enacting plans	(C3) Following rules of engagement, (e.g. prompting other team members to perform their tasks)
(D) Monitoring and reflecting	(D1) Monitoring and repairing the shared understanding	(D2) Monitoring results of actions and evaluating success in solving the problem	(D3) Monitoring, providing feedback and adapting the team organisation and roles

Table 3: Matrix of collaborative problem-solving skills for PISA 2015 (Source: OECD, PISA 2015 Assessment and Analytical Framework, 2017, p.137)

An example for the PISA 2015 CoIPS assessment tool can be seen in Figure 5. The interface shows a restricted amount of interactions such as the selection of possible answers given to the participant instead of using their own language which resembles an enhanced multiple-choice question.

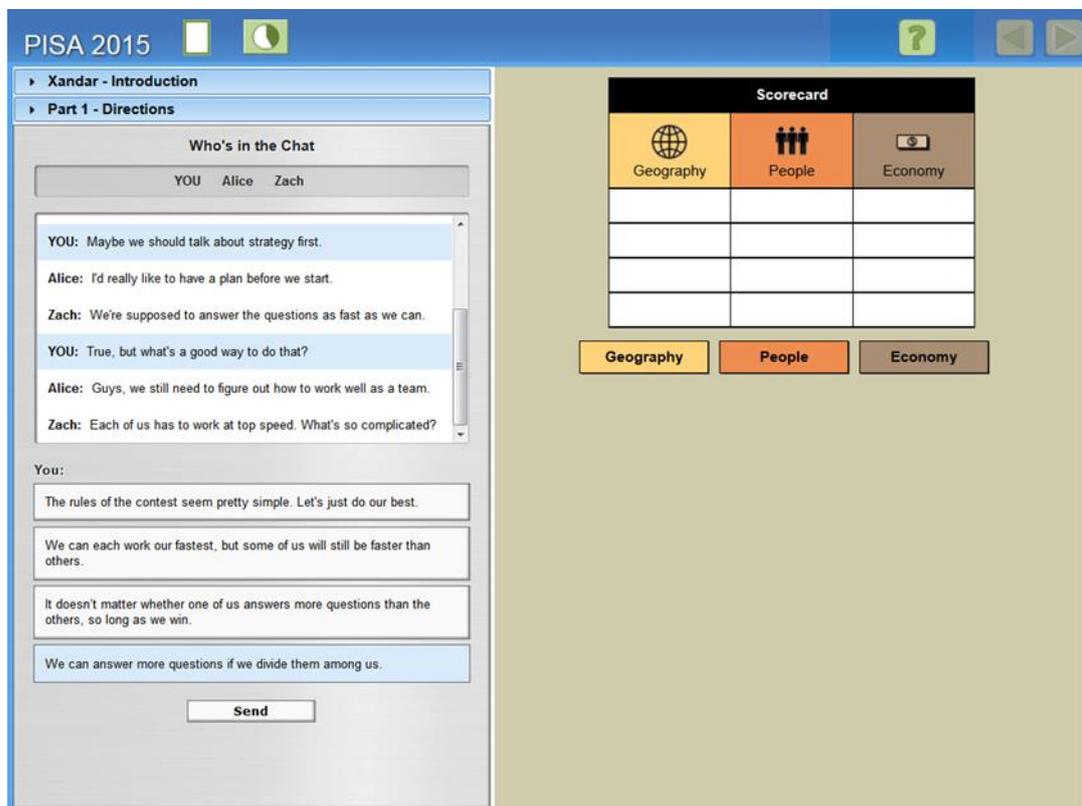


Figure 5: OECD Xandar Item, "Part 1, Item 4: Agreeing on Strategy" (Description of the released unit from 2015 PISA collaborative problem-solving assessment, OECD 2017, p.3)

Besides OECD with its PISA study, research about adequate test setups in ColPS assessments approach two major directions of interaction: human-human interaction assessments, where test-participants solve problems together directly, and human-computer interaction assessments, where the test participant interacts with a computer-based character to solve a problem.

Both approaches have advantages and disadvantages when it comes to validity of the results, standardization of the items, and control of the test setup. According to Greiff, Holt and Funke (2013, p.83ff), the human-human approach of ColPS assessment is high in face validity but is very hard to control. On the other side, human-agent approaches of ColPS assessment support a good level of standardization and enable provision of a controlled testing environment. However, after OECD's PISA 2015 assessment of ColPS skills, it became clear that the human-agent assessment maintains control and standardization but prohibits flexibility and reduces real-world reference and thus comes with a lower content validity.

There are many different setups and field trials when it comes to the examination of group behavior in problem solving. The ATC21project led by Griffin and Care (2012) had a human-human setup where a group of two was communicating through a chat-interface. The biggest limitation that has come to light is the exclusion of automatic scoring when both agents (students) make use of natural language. Rosen and Tager also approached the assessment on the individual level (2013), attempting to offer open space with natural language as a human-human setup and the interactive task solved in dyads. Another approach can be found in the research of Cukurova et al. (2016) who developed a framework for collaborative problem-solving in practice-based learning activities. The developed environment follows the PISA ColPS framework, but is enhanced by several factors that include awareness of a natural situation of collaboration. The Learning Analytics System collects data from both ambient and live sources while the learning environment is open and designed to support collaboration. However, this approach has limitations. The data collected through this open space is meant to be rich and contribute to the field of ColPS research, but it is not suitable as an assessment tool since the definition of scoring events in a semantic context has not yet been solved. Hao et al. (2015) emphasize the challenges of developing a psychometrically rigorous ColPS assessment; numerous complex factors need to be considered and include the type of task or the skills and the personalities of the team members.

## 2.2 Conversational Agents

### • Background

In order to understand what makes a “conversational agent”, both elements of the term must be defined. *Conversation* according to the Cambridge Dictionary is “*talk between two or more people in which thoughts, feelings, and ideas are expressed, questions are asked and answered, or news and information is exchanged.*” This definition implies that a conversation is initiated by two or more parties, and Radlinski and Craswell (2017) call this a “mixed initiative”. The term *agent* can be defined in a variety of ways. Franklin and Graesser (2005, p.22) compared several definitions, two of which fit best with the context of this thesis:

**The IBM Agent:** “Intelligent agents are software entities that carry out some set of operations on behalf of a user or another program with some degree of independence or autonomy, and in so doing, employ some knowledge or representation of the user's goals or desires.”<sup>1</sup>

**The Wooldridge-Jennings Agent:** “... a hardware or (more usually) software-based computer system that enjoys the following properties:

- **autonomy:** agents operate without the direct intervention of humans or others, and have some kind of control over their actions and internal state;
- **social ability:** agents interact with other agents (and possibly humans) via some kind of agent-communication language (Genesereth & Ketchpel, 1994);
- **reactivity:** agents perceive their environment, (which may be the physical world, a user via a graphical user interface, a collection of other agents, the internet, or perhaps all of these combined), and respond in a timely fashion to changes that occur in it;
- **pro-activeness:** agents do not simply act in response to their environment, they are able to exhibit goal-directed behavior by taking the initiative.” (Wooldridge & Jennings, 1995)

These two definitions make clear that agents represent an interactive, artificial element or program that is interacting with a human being in a certain manner. This theme combines the elements of the conversational agent logically.

### • Agents' System and Technology

Conversational agents have been around for some time and have gained considerable attention in recent years. By the 1960s, natural language was simulated with text-based dialogue systems. According to Radlinski and Craswell (2017), the definition of a conversational system is an information retrieval system that permits a mixed-initiative between an

---

<sup>1</sup> IBM's Intelligent Agent Strategy white paper (original source not available anymore)

agent and user. Here, the agent's actions are based on the conversation, using both short- and long-term knowledge of the user. According to this definition, a conversational system needs to have at least five properties:

- **User Revelation** - the system helps the user to express their needs.
- **System Revelation** - the system is clear with its capabilities to form user expectation of the system.
- **Mixed Initiative** - both system and user can take initiative for conversation.
- **Memory** - the user can reference past statements and the system understands.
- **Set Retrieval** - The system can reason about the utility of sets of complementary items.

These properties outline the necessary interaction given by a conversational agent. Another part of the conversational system that needs further definition is the processing of the language itself. Within the field of NLP and machine learning, "chatbots" are often created based on structuring dialogues with AIML (Artificial Intelligence Markup Language). AIML offers an XML-based approach into the less technical structure of natural language, initiated by Richard S. Wallace who is describing it as follows: "*The primary design goal of the original AIML language was simplicity. AIML is motivated by two observations:*

- *Creating an original, believable chatbot character requires writing a significant amount of content, in the form of conversational replies. [...]*
- *The people who are most suited to writing the bot content are not, by in large [sic], computer programmers. Those with literary backgrounds are more skilled at developing content for original characters."* (Wallace, 2014)<sup>2</sup>

Furthermore, some agents have AIML implemented. AIML is also intended to be a system and method for atomized creation of chatbot content from scripts of conversations. Today's technology has made huge advances, and solutions are available to create conversational agents without extensive and complex dialog structures in markup language. The declared goal of AIML that Wallace (2014) describes remains relevant, but AIML has been superseded by new technologies. Most conversation-based systems now use JSON as the language to structure conversation flows in a machine-readable manner, but the system behind the language processing closely resembles AIML technology.

Typically, the architecture of a conversational agent consists of three main components. The graphical user interface (GUI), the chatbot architecture subdivided into NLU component, processing unit and the database. Most agent-based systems are enriched by several connected APIs.

---

<sup>2</sup><https://gist.github.com/onlurking/f6431e672cfa202c09a7c7cf92ac8a8b> (retrieved 20.06.2018)

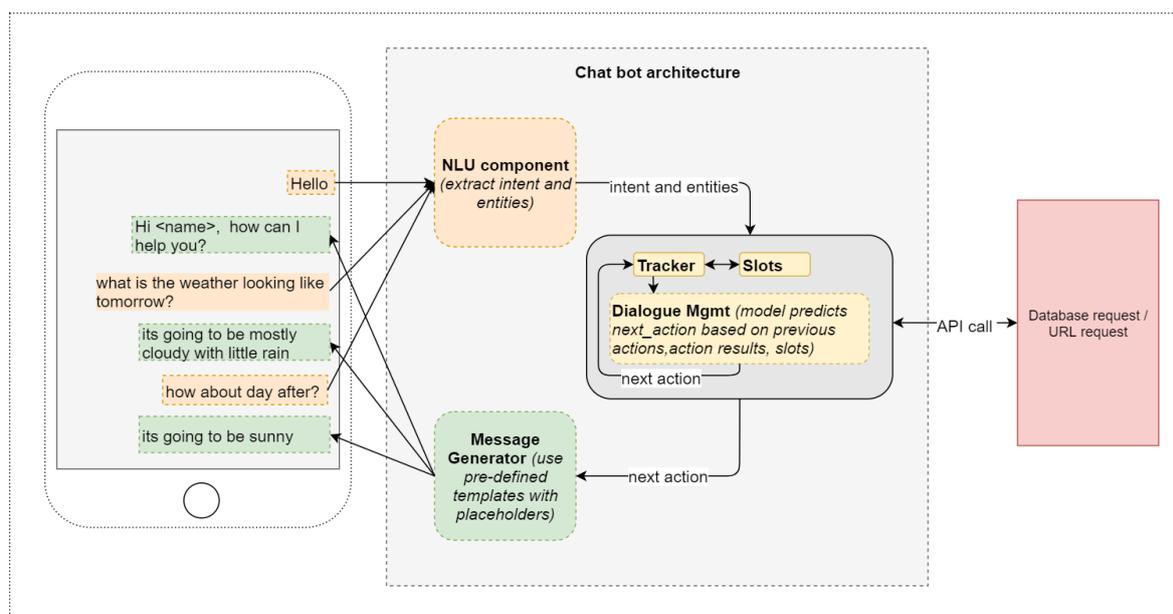


Figure 6: Chatbot Architecture of a conversational AI chatbot.  
(Kompella, 2018 in: Towards Data Science)

The importance of the examination of this topic is shown by the increasing appearance of chatbots in our daily lives. Today, chatbots are used to clarify different tasks across various businesses, including e-commerce, insurance, banking, healthcare, finance, legal, communications, logistics, retail, automotive, leisure, travel, sports, entertainment, and media (Davydova, 2017). Since these industries would not be able to use this widespread technology without science, it is reasonable that scientific fields inherit “chatbots” or conversational agents to widen research and development.

### 2.3 Synthesis of the Literature Analysis

Following Blessing and Chakrabarti’s Design Research Methodology (2009), the literature analysis aims to define the aim, focus and scope of the research to be conducted. Considerations that came to light with this literature review can therefore be used to define this research paper. The aim of this project is to create a human-agent environment that follows Radinski’s and Craswell’s (2017) definition of a conversational system that serves as a conversational system to simulate a collaborative problem-solving scenario. The focus of this thesis is to unveil useful enhancements of a human-agent system. This focus will be given through an empirical study, presented in Chapter 3. The scope of this thesis is to develop five items of the Xandar unit that was published in PISA’s CoIPS assessment in 2015. Equally, it should suggest approaches to the dimension of dialogues with collaborators in the assessment as well as a suggestion to the problem-solving space.



## 3 Empirical Analysis: Expert Interviews

### 3.1 Methodological Approach

Bogner and Menz (2009, p.46) define the expert interview as an instrument that has its origin in empirical social science research. The goal is to collect data that can be used as an exploratory tool. The main focus of the interviews in this thesis is the clarification of questions about ColPS and its current operationalization. The goal was to conduct interviews with at least five research experts of the field, with the aim of answering general questions as well as technology-focused questions in order to benefit from their expertise. The implementation of the expert interviews follows a problem-oriented approach with open interview questions as described by Mayring (2002, p.67) and Bogner and Menz (2009, p.46).

### 3.2 Operationalization

The literature review in the previous chapter gives insights into the current state of research on ColPS and ABS. This background helps in understanding the role of ABS within the ColPS field and informs the division of the wider topic into the following different dimensions for the development of expert interviews:

- **General questions about ColPS Assessments:** The goal is to find out about urgent needs or gaps in current ColPS research, according to experts who are working in this area.
- **Matters of understanding and creating artificial characters:** This question aims to understand the researchers' expectations of an artificial character and where there is need for improvement.
- **Matters of understanding and structuring communication in ColPS:** The question aims to understand the researchers' expectations towards communication within ColPS assessment.
- **Matters of understanding the assessment instrument:** The question aims to understand how effective current technology is (e.g. CBA ItemBuilder) in creating ColPS items or to examine retrieved data.

For each of these dimensions, subordinated research questions (open and half-open) have been defined. Three to six in-depth questions were created, within a well-structured interview guideline. After the first interview, some of the questions have been generalized in order to be able to conduct more focused questions and to reduce repetitive statements. The interview guidelines can be seen in Appendix A.1. The shaping of the questions has been done in such a way that experts who did not participate in the PISA 2015 assessment could also answer. The interviews were to be conducted within 45-60 minutes. The last paragraph of the interview guideline includes sociodemographic questions and information about the

research context of the expert. This data is not included with the transcriptions in order to maintain the privacy of the interviewees.

### **3.2.1 Sampling**

*“In a pragmatic perspective – focusing on the local context of knowledge production, the status of expert could be understood as ascribed by the researcher: a person is attributed as expert by virtue of his role as informant.”* (Walter, 1994, p.271) Therefore, the experts have been chosen to be researchers in the wider field of ColPS. Experts that suite this context were found through their research and were publishing their own research experience and experiments in the ColPS field or were contributing to PISA 2015 directly.

### **3.2.2 Experts**

This section gives insights to the background of the interviewed experts and their level of proficiency.

- Expert A: Associate Professor at a university, educated with a psychological background. Expertise is mostly within the cognitive part and the assessment part in the ColPS field. The expert is part of a team of around 15 people. Further research within the field of ColPS is planned.
- Expert B: Post-PhD researcher and assistant lecturer with an educational background. Expertise is in human-human and human-agent setups of the assessment with a focus on the collaborative aspects of ColPS. The team size varies with the projects and funding. A shift of research focus is possible depending on funding.
- Expert C: Professor at a university, focusing on cognitive science, artificial intelligence and educational software. The team size varies with projects and funding from 15 up to 100 people including teams at partner institutes. Research within the field of collaboration and life-long learning is planned.
- Expert D: Post-PhD researcher at an international institute with an educational background. Expertise is varied but is focused on psychometrics. Team size changes but is typically about 4 members. ColPS is not the main research focus but will be part of future research interest.
- Expert E: PhD senior lecturer with a background in computer science. Expertise in ubiquitous environments and collaboration to foster formal and informal inquiry-based learning. The team size depends on funding and projects but is between 4 and 10 members. Research focus within ColPS will be maintained in the future.

### **3.2.3 Data Processing and Analysis**

The interviews were conducted via Skype or, in case of connectivity issues within the country of the interviewees, through a technology of their choice. The interviews were recorded on the PC of the interviewer, as well as with a second external device as fallback. Five

interviews were conducted, although the recording of one interview was corrupted through technology issues. Since notes were taken during this interview, these will be included in the analysis. The interviews were conducted in English only, in order to reduce distortion of the results through subsequent translations. The audio files have been transcribed into a minimal transcript (Selting et al., 1998, p.8ff). Transcripts were examined jointly so they became clear to any recipient. Also, the adjusted interview transcripts offered a better basis for analysis according to the qualitative content analysis approach of Mayring (2010, cited in Mey G., Mruck K. (eds) *Handbuch Qualitative Forschung in der Psychologie*, p.601-613). All transcriptions can be found in Appendix A.2 of this thesis.

As mentioned above, the data that was gathered through the interviews was then analyzed following the procedure of Mayring's qualitative content analysis in an abridged form. The results were transmitted into a Microsoft Excel table and can be found in Appendix A.3. The data was set out in a matrix that supported allocation of results given by the interviewees in order to perform a better analysis. This approach made it possible to carry out the evaluation of the interviews in accordance with the theoretically derived topics (see Mayring 2010, p.57ff).

In a first step, all the interview transcripts were read, and possible structures and patterns were noted. In the next step, the answers of the experts were structured, analyzed and evaluated according to the selection criteria given by the question categories. Depending on the research question and material, the appropriate analytical technique (summary, explication, structuring) was chosen (Mayring 2010, p.65).

Newly-discovered insights that were gained through this technique then led to further development of hypotheses and research goals for this thesis and could conceivably lead to further research.

### **3.3 Analysis of the Expert Interview**

#### **3.3.1 Description of Results**

The results to the enumerated dimensions given in the previous subchapter were evaluated with the *Qualitative Content Analysis according to Mayring* in mind. In this section, only interview data is presented and it is not related to the theory. A conclusive discussion of the interviews will be presented in the next chapter.

#### **General Questions about ColPS Assessments**

- *Do you feel confident with the current state of ColPS Assessment with agent-based systems (e.g. such as has been used in PISA)?*

There was not a consistent level of confidence in the current state of collaborative problem-solving. The novelty of the entire field, especially in the context of digital assessment, was emphasized by all participants of the interview. Two researchers who contributed to PISA 2015 commented that in the context of the assessment, the approach that was followed was the best possible (Interview 1 and 3). The limitations that come with the preservation of the validity of an assessment enforce restrictive conditions on the CoIPS construct (Interview 1 and 2). This is why two experts mention that they were not fully satisfied and one was not at all satisfied with the current state of the assessment (Interview 2, 4 and 5). Reasons for satisfaction were the high quality of standardization and the resulting feasibility of a large-scale assessment like PISA (Interview 1). Objections to this view centered on the lack of enrichment for specific situations in the context of problem-solving such as negotiations. Another point of criticism was that the sample size for such a young technological construct has been over dimensional and issued to the public without sufficient testing beforehand (Interview 2 and 4). Another gap in the assessment that was emphasized was the importance of the scoring of an individual's performance, although another expert presented a contrary view, saying that the performance of the group is important (Interview 5). This point is in reference to the situation where an individual might be rather introverted in oral contributions but is good in performing other tasks that contribute to a solution. One expert emphasized that dissatisfaction arises due to the lack of completeness and awareness of research that was conducted before the PISA 2015 assessment. All interview partners agree that the extent of the construct itself has not been captured in an ideal way or in its entirety.

- *Do you feel confident with the current state of technology available to conduct collaborative problem solving with students?*

There was agreement between the interview partners in questioning confidence in the state of technology. Three out of five participants mentioned the imitation of natural collaboration in several contexts (Interview 1, 4 and 5). The agreement is around the point that emerging technologies open the space for further research on increasing the naturalness of collaboration such as use of natural language. In the context of PISA, one expert stated the importance of a differentiation between what is possible and what is useful in the context of this further technical development (Interview 1). The technical background of the team behind the PISA 2015 assessment is emphasized as versatile in all dimensions (technical, psychological, educational), yet the feasibility of the imitation of a natural collaboration with agents has to be questioned (Interview 1, 2 and 4). In contrast, another interview partner believed that exhausting the rich state of technology could definitely result in sufficient authenticity

for an assessment (Interview 5). The same person stated that the authenticity of collaboration scenarios is not achieved in the assessment and, linking back to the question, semantic extraction is limiting development. Three out of the four interview partners agreed on this limitation and mentioned the major challenge of PISA, which is the multilingual aspect of this specific international assessment. The scale of the semantic complexity, especially for multilingual assessments, is the one restriction that recurs throughout the interviews.

- *Other than CBA ItemBuilder, do you know, or have you worked with other tools to create Items for collaborative problem solving? If yes, what did you like or dislike?*

Four out of the five experts are carrying out research with their own technological artefacts in the field of ColPS, and two of these artefacts are also able to handle assessments (Interview 2 – 5). The expert in the first interview did not specifically state which platform was being used. It was emphasized that the freedom of educators to use a technology of their choice in order to create valid course or assessment designs should be a given. Yet, the availability of tools is rather limiting educators in the creation of their lecturing or assessment artefacts.

Three out of five interview partners were familiar with CBA ItemBuilder, which was used for PISA 2015 assessment (Interview 1, 2 and 4). Those who were familiar with the platform agreed on the objective fact that the platform available for a large-scale assessment was building a solid base for the context with an increasing complexity. Nevertheless, the difficulty of considering all aspects of an assessment of this scale is a major factor. The absence of software that can fully meet needs in the educational context is mentioned to be a universal issue in the field, and one expert (Interview 5) emphasized the huge difference of technology available to the educational sector as against, for example, the gaming industry. Likewise, another expert expressed concern about the non-existence of assessment platforms that could incorporate technology such as augmented or virtual reality (Interview 4).

- *If you could name the most important aspect of the assessment of collaborative problem solving that needs improvement - what would it be?*

Opinions on this question were manifold, and there was no major consensus to be discovered amongst the interview partners. As a technical improvement, interaction itself was mentioned to be improved. This means an increase of the variety of interactions possible to the test taker or simply something different from multiple choice answers (Interview 1 and 5). The system should also be able to react better to the student's interaction or offer follow-up interactions (Interview 3). One person emphasized the lack of general background research to obtain a better general

understanding of the nature of collaboration (Interview 2). This undercurrent can be heard within all the interviews and also in other questions: In the field of ColPS assessments there is no solution in which the nature of collaboration is mirrored commensurately.

### ***Matters of understanding and creating artificial characters***

- *Can you briefly describe the interaction of a student with an agent in a current ColPS assessment?*

All interview partners were aware of the PISA 2015 assessment setup. It mostly has been described as a rather limited interaction that simulated a chat-based environment but set up as a multiple-choice assessment. The reason for this closed interaction was the standardization that is necessary for such a large-scale assessment. One researcher described their own research approach compared to the PISA assessment, where the setup of the test environment is a human-human interaction (Interview 2). With this approach, the researcher initially left the conversation open, allowing natural language. This unveiled the complexity of natural language implementation to the assessment instrument. Obtained data of this study could not be processed appropriately, and there was no possibility to score the conversational transcripts. This experience brought the research team to another human-human approach, where the choices a test participant could make were pre-defined messages. The reason for the renunciation of an agent was to obtain a more realistic situation of the collaborative problem-solving process.

- *How would you ideally expect an interaction with an artificial character in an agent-based system?*

This question might appear redundant because it is the inverted formulation to the previous question. Yet, it demonstrates its value by obtaining profound results and to reduce cognitive dissonance of the interview partners. To retrieve the expert's expectation of artificial characters in ColPS, this inverted questioning is specifically posed and brought manifold answers. Even though every expert had a clear idea about possible enrichments, two out of four interviewees said that a profound definition of a best possible interaction is hard to define (Interview 1 and 2). Also, two experts could state that the authenticity of the agent should be enriched and that they should simulate a more natural behavior of collaboration (Interview 1 and 4). Although the other participants have not directly formulated the general extension of authenticity, they have steered in the same direction with their other statements: the agent should be able to obtain and process the student's behavior and adjust its reaction to it. Another suggestion follows the same direction, stating that the response

to the kind of interaction a student performs should be more versatile, for example when the student needs more iterations to find a good solution (Interview 5). This includes the very basic expectation that the agent understands what a student is saying. All experts were addressing emotional aspects of an agent as an ideal enhancement.

- *If you have to build CoIPS Items now, is there anything else missing for the best possible implementation of artificial characters for CoIPS Items?*

To narrow down the best possible outcome of an agent that could be built, this question aims to gather information about other missing implementations. Answers that were given emphasized the desire for more naturalness in the CoIPS assessment situation. This opinion was shared by four of the five participants with a high priority (Interview 1 to 4). It includes the increase of the authenticity of the collaboration itself, but also of the avatars or artificial characters. As one expert is mentioning, the privacy and the ethics of such a situation should be considered (Interview 5). The ethical part is nearly not considered in any publication that discusses the CoIPS construct. However, it certainly deserves more attention to be prudent about the preservation of data and the user's privacy.

After all, one expert is agreeing to the necessity of an increased richness of different patterns for interaction. This richness should be applied to the artificial character, such as a sort of a physical interaction or voice implementation (Interview 5). Yet, these wishes tend to defeat the purpose of the assessment.

- *How would you describe the perception of the student through the system? Does the student's behavior (anger, mistrust, fear or doubt) influence the artificial characters? Is there a channel to recognize the student's reaction (especially when messages are pre-defined)?*

This question is focusing on the system's perception of the student's interaction with it. Four interview partners state that there is no perception of the student through the system. Though the system of PISA, CoIPS is reacting in a certain way to the response that was given; the reinforcement of the student's reaction is stated as imperfect. One expert explains this by the minimalist setup of the CoIPS items with predefined messages where the scoring is based on the choices selected. This emphasizes the focus on the cognitive part of CoIPS and not on the social part of a collaborative situation. A study that is mentioned many times by the experts throughout the entire interview, but also in the context of this question, is research conducted by ACTnext. This research on CoIPS is collecting a lot of information about the test participants like, for example, the position in the room, and then the system adapts to these kinds of information. Two interview partners took the opportunity in

this question to give suggestions or further considerations about these circumstances (Interview 1 and 5). One is to do further analyzation on the implementation of natural language and how this could be implemented to maintain the validity. One other expert stated that it has to be considered that students do not show a major interest in educational software in general, why it should be scrutinized if it is necessary to enhance an assessment to this extent, or if it wouldn't be sufficient to just focus on a series of submissions for example (Interview 5).

- *How does the artificial character influence the quality of the item or are there other aspects to be considered?*

This question tries to get to the bottom of quality assurance of an assessment, when having implemented artificial characters. The general tone of experts in this interview is that agents influence the validity of the assessment. They have a major impact on the assessment since people's reaction to artificial characters can be versatile. One expert is stating that more agents will leave less space for the human to interact (Interview 2), while this could reveal a contradiction: another expert states the more complex a problem is, the better it can be solved with more participants (Interview 5). Another consideration that was mentioned by two of these interview partners is the consideration of the identity of the character and the impact that it can have. If an artificial character is of a certain heritage or gender, or is speaking with a higher or lower voice, this might influence the student's behavior in a collaborative situation based on personal convictions of the student and their own social background. One interviewer is emphasizing the difficulty of this question where there might not be a correct answer yet without further empirical research.

- *Do you think a personalization (as in personification) of the artificial characters help to enrich the item type?*
  - *visual personification (e.g. providing faces with mimics)*
  - *linguistic personification (e.g. talking in a rather informal language)*
  - *characteristic personification (e.g. strong characteristics)*

Ultimately the dimension of the personalization of an artificial character is of interest for this thesis. Three of the interviewed experts agree on a certain impact of characteristics to the student and therefore to the assessment. Yet, again it is mentioned that this question would need further empirical research (Interview 4). The dimensions of influence that the implementation of characteristics could have are versatile. Two interviewees agree that it would enrich the entire experience of the assessment and might have an impact on motivation of the participants (Interview 1 and 2). One very concrete and meaningful statement is that at this moment in PISA

2015, the choice of an answer will not initiate a reaction to the agent: *“As long as this is missing, also the other improvements do not have a bigger impact.”* (Interview 2). Hereby, one expert was differentiating more on the suggested properties of an agent. The interviewee concludes that a visual enrichment could increase the engagement of the student, but also the noise of the item (Interview 1). The personification is an influence that should be implemented with the construct of collaboration in mind.

Furthermore, two out of the five experts emphasize the cultural background that the agent could mimic, as well as the representation of the agents by gender, voice, age and others (Interview 4 and 5).

### ***Matters of understanding and structuring communication***

- *Can you briefly describe the communication, taking part in a typical ColPS assessment?*

The first question in this section focuses on the procedure of communication that is taking place in a ColPS assessment. Three of the experts were referring to the PISA 2015 assessment (Interview 1, 3 and 4), while two were answering with their own research projects in mind. The communication implemented in PISA 2015 is described as a system reacting to a student's choice of answers given. One expert is describing the situation the most concrete: The student gets introduced to two agents and gets presented a selection of answers (Interview 3). This makes it a multiple-choice interaction. Since the number of choices is discreet, the agent's reaction can be irrespective to the student's reaction. This is why the reinforcement of a student's reaction is considered as imperfect. Another argument that came to light is that the focus within the PISA assessment is on the topic itself, whereas it doesn't allow any off-topic conversation. This is namely the biggest difference towards a real-world situation (Interview 4).

One expert is describing the communication in the own conducted research as restricted, where the students can send predefined messages, images of the screen or diagrams (Interview 2). This decision was made to maintain the validity of the assessment, due to high complexity that a natural communication brings. One other expert is trying to implement communication as it happens in daily life into his research. He discovers that communication varies with the engagement of the students who are collaborating. Some might be distracted, and the participation might vary in general. For example, some students are less physically involved, but share a high amount of information, while others are not participating to the communication but work silently on concrete solutions to a part of the problem (Interview 5).

- *How would you ideally expect communication to happen in an agent-based system?*

Four experts were giving answers to the question how the communication should ideally be in an agent-based system. The overall agreement on the free and open integration of linguistics is present and considered with different arguments.

A consideration prohibiting the implementation of free speech of a student is the lack of instruments to analyze the semantics of the statements. Other factors that can be considered are the implementation of time-based interactions. The interview partner refers this suggestion to the current state of CoIPS assessment in PISA 2015, where a student has unlimited time available, but is scored by the result. This isn't coincident with the natural situation of problem-solving in our daily life, where a contributor to a problem-solving situation much rather has a certain amount of time available and can approach with several attempts to find a solution to the problem within the given time. A similar direction can be heard from another participant who emphasizes that more back and forth interaction would increase authenticity. Another estimation of a best-case scenario was presented, where also inter- and intra-group interactions are considered. Restricting analysis to the dialog only is not realistic because a collaborative problem-solving scenario is just as limited to text-based interactions only.

- *If you have to build a CoIPS item now, is there anything missing for the best possible solution for communication within a CPS assessment?*

Again, participants are jointly agreeing the necessity of the enhancement of the interaction in several ways. Two participants are comparing the state of current CoIPS assessments to the state of the art in the gaming industry and point at the gap in technological realization. The recent approach as a step-by-step approach is criticized to be too discrete, and the integration of a continuous flow of interaction is proposed. The same two participants give suggestions on the enhancement of interaction through the integration of talk or small talk, allowance of search processes, or similar situations as they are given in classical gaming. Also, the tracking of conversation flows is mentioned, comparable to office situations where grapevine is common. In conclusion, there are two statements that bring up critical considerations; one is the collection of rather less data, but the right one and the other suggests that the most realistic assessment of a CoIPS situation would be to assess one person with a thousand other persons in a problem-solving situation to obtain a quantifiable meaning, which also is not realistic.

***Matters of understanding the assessment instrument***

- *Is there something missing when you think about currently used agent-based system, that a real-life interaction has, but the current assessment instruments do not have? If yes, what is it?*

This question is posed to understand given restrictions by the instrument. One person is referring to the restrictions that the ColPS framework brings to the assessment. For the assessment itself, enrichment through video or voice recording is one aspect to be considered. This statement goes with another expert's opinion where the consideration of the macro space such as the entire classroom should be made. Furthermore, a desire for in depth enhancement of social interaction has come to present with this question. The assessment of feelings and descriptors of the human-human discussion are wished at the side of the test-taker. While the implementation of a simulation of emotions, facial expressions, an interaction between the agents or in general more reactivity are considered as missing parts when it comes to the enhancements of the assessment instrument. On the rather technical side, there is mentioned that even the implementation of pauses or delays would bring an enhancement to the instrument because it would increase the authenticity of the conversation.

- *Did you have to process the obtained data that a ColPS (e.g. CBA Item Builder) Item generates (as in quantitative research) and how practicable was it for you to handle the data in order to conduct quantitative research?*

Two out of five interview partners were familiar with the data of the PISA 2015 assessment. The data processed for this assessment are extensive and mainly expert oriented. This can bring certain constraints but works well if the researcher has a certain goal in mind. The other three participants were able to tell from their own research constructs how data structures are set together within this research field. One researcher was missing specific data in terms of actual meta or para data that could be given by the system but were not preserved or not published throughout the study. One certain restriction is the lack of clear standards for assessment data. In terms of the item-data itself, one expert highlighted its sparseness; thus, it was only saving binary data, giving information whether the student was clicking the element within the item-section or not. One researcher described problems that generally come with assessment data or data from sociological studies, which is the amount of data to be processed. This constraint will be resolved as soon as richer technologies such as unsupervised deep learning evolve and can be used for automatic scoring.

One person described data that is provided by their own developed system and emphasized the complexity of collaborative problem-solving data. It is important to

maintain the ability to manipulate the data base in order to get specific results. The consensus amongst the interview partners was given that fortified data would be an enrichment to the construct.

- *Where do you see agent-based systems in your daily life? If you have seen some or are actually using them, which ones are you using? Which ones do you like the best and what do you like about them?*

In order to understand whether the interview partners tend to rather reject or accept these kinds of technologies in general, this question focuses on the researcher's attitude. Three researchers were describing that they either have agent-based systems in their daily life, or as a part of their research (Interview 3 - 5). Two of them are using the virtual assistants that come with most operating systems now. These are, for example, Siri by Apple, Cortana by Microsoft, or Alexa by Amazon. One researcher was mentioning the work with own agents in conducted research, who serve their user by advising them as tutors or as lifelong learning assistants (Interview 3).

### **3.3.2 Discussion of the Results**

The results that emerged in the interviews will be discussed in this chapter associated with appropriate theory. Results should be put in a context with current discussions and standards.

#### **General Questions about CoIPS Assessments**

This section of the interview aims to gain insight into the general attitude of the experts towards CoIPS. Interview partners argued about the general research background to the PISA 2015 assessment. Krkovic, Greiff, Pásztor-Kovács and Molnár were discussing the constraints of the assessment of collaborative problem solving in their publication in 2014. In this publication they were able to unveil major gaps and problems in the existing attempts to provide assessments of collaborative problem-solving. This underlines the tone of interview-partners. Discussed weaknesses of the assessment could have served as an indicator of necessary further research in advance of the large-scale assessment.

Hence, one expert is mentioning the necessity of reducing the reading load. This argument is being reflected in considerations, published in Japan after the PISA 2015 Assessment. The results of this study on secondary data of the assessment (Komasu & Rappleye, 2017, p.619) were confirming the concerns of policymakers and mainstream media in eastern Asia. The introduction of computer-based testing had the effect of lower reading scores in that region. This consideration does not apply to the CoIPS assessment, which specifically emphasizes that computer-based assessments can have an impact on the construct of an assessment in general. Komasu and Rappleye are stating that the most important argument

about their results is that the shift to computer-based assessment in PISA 2015 may have invalidated the OECD's attempts to provide comparable longitudinal data to the world. Contrary to this, Greiff, Holt and Funke were stating the advantages of Computer-Based Testing in 2013 (p.85). This setup provides a basis for a controlled approach to collaborative problem-solving and opens the field to a more manifold data analysis ex ante. This argument also can be connected to the argument about difficulties in maintaining content validity and measurability of an assessment, which was brought up by several experts in the interview as well. Nevertheless, further research on the impact of reading load within computer-based assessment might be a direction the field can profit from. This introductory part of the interview has already revealed some powerful arguments to the deficits that ColPS constructs in a digital environment still have.

### ***Matters of understanding and creating artificial characters***

The background of this question reaches out to OECD, saying *“PISA took advantage of computer-based delivery and let students interact with computer simulations of humans (known as computer agents) whose behaviour can be controlled. [...] Students’ performance with the computer agent was a moderately good predictor of their performance with the human partner.”* (OECD, 2017).

The current state of the artificial characters implemented to the assessment has been described with a general critical tone by the experts. The publication of the PISA 2015 results, but much rather the considerations and the collaborative problem-solving framework were opening the field to further research on better approaches. Yet, there is rather little literature about the generation of artificial characters within ColPS constructs. Participants in the interview were finding fault in the variety of influences of the assessment, that an artificial character could have. These are namely; the tracking of emotional processes, having emotional cues like avatars or facial expressions, or simply and generically, a “human side”. Early agent-based systems were simulating a human performance of simple tasks by creating goal-oriented and data-determined behavior (Hung, Elvir, Gonzales & DeMara, 2009, p.1236). With the improvement of technology through several developments such as Automatic Speech Recognition, the necessity of more natural dialogs arises. Hung et al. argue that, currently, the spoken interaction may not be as efficient as accomplishing tasks as text-based interaction. Yet, a factor that certainly can have an impact is the length of a conversation. According to Schumaker, a conversation length is an important metric in maintaining dialog quality (Schumaker, 2006). Targeting the interview-partners’ desire or even previous research experience with avatar-based interfaces, literature confirms the positive impact of avatars within an environment (DeMara, Gonzales, Hung, Leon-Barth, Dookhoo, Jones, Johnson, Leigh, Renambot, Lee and Carlson, 2008). These avatar-based

interfaces can reduce the user's machine-interface workload and allow the user to more completely focus on the task itself. The resolution of multiple-choice type assessments can be seen as a certainty, if one is comparing current technologies published to the market. Thereby, the interviewees have repeatedly referred to the games industry, which is an environment that is well known to young participants. The gap between educational sector and the gaming industry can only be estimated, but according to experts in this interview, the educational sector has not the power to improve at the pace the gaming industry has done due to a lack of monetary resources and complicated founding systems. Further effects and implementations to the artificial characters that a participant should face but are not implemented at the current state are effects of time pressure, conflicts, stress or peace in an environment, working with competent or incompetent people (Krkovic et. al, 2014). Certainly, the research within agent-based environments is much richer, as the literature review already has shown. Also, we all have passed some time with ELIZA in the past. This is why the integration of conversational agents into the ColPS construct seems like a promising approach. Interview partners were agreeing on this.

### ***Matters of understanding and structuring communication***

The questions in this part of the interview were aimed at understanding the researchers' expectations towards communication within ColPS assessment. A major part of collaboration is the communication between team members, and this leads to the need to ask experts about the key matters of communication structure when creating a ColPS item.

According to Krkovic et al. (2014), "one way to structure communication in a collaborative problem-solving task is to use chat boxes." The interview results emphasized that this is not the best practice that could be applied. The currently implemented structure of communication in most ColPS assessments is restricted, and this leads back to the unsolved problem of semantic analysis. This is a complex field in itself and it brings major problems when it comes to automated scoring. Currently, there is no solution to the problem of automated scoring through semantic analysis, and this is keeping the assessment far from the implementation of natural language. One expert described the different roles a participant can have, which are currently not considered, at least not in within the Xandar Item which will also serve as the basis for the further evaluation of this work.

Several publications make the point that awareness of different group compositions is omnipresent in the field (Fall et al. 1997; Rosan and Rimor, 2000, Webb, 1995; Wildman et al., 2012). It is understood that students show differences in contributions and may be rather quiet but productive or producing lots of noise but not contributing to solve the problem. The same effect applies to the intragroup relationship, where students might perform better or worse when being exposed to different team members. Another aspect that came to light in

the interview is the consideration about implementing emoticons. Derks, Bos and Von Grumbkow (2008, p.99ff) conducted a study in which the role of emoticons in computer-mediated communication was observed. Their results showed that emoticons are mostly used to express emotions, to strengthen messages or to express humor. Furthermore, the participants in this study seemed to use emoticons that were similar to a facial expression that would emphasize their relating statement.

Certainly, this section brought important aspects to light when it comes to the implementation of communication into ColPS Items. Most aspects are of a sociopsychological nature, such as the impact of different characters or the impact of the usage of emoticons. These considerations will influence the artefact that is being presented with this thesis.

### ***Matters of understanding the assessment instrument***

Research on collaborative problem-solving does not solely focus on the assessment, but also on the nature of the scenario itself. However, most research is conducted within the field of the assessment of collaborative problem-solving. The last paragraph of the interview focused on understanding the assessment instrument and examined how adequate the current state of technology is in order to create and enrich ColPS items or to examine its produced data.

Interviewees mostly commented on missing implementations that refer to the item content and suggested only minor improvements to the available technology. This point leads back to the fact that most of the experts who participated had only limited backgrounds in the technical field, or it might be that they consider it more important to enhance the assessment with content-focused features. The results are varied, and again it becomes clear that experts are formulating a wish to integrate open-ended computational linguistics. Another factor that came up in the previous section was the implementation of pauses, delays or integrating not only inter-group interaction but also intra-group interaction. Rosen and Tager (2013) and OECD (2013) were likely to be aware of the missing variety in possibilities in ColPS assessing environments. Nevertheless, they state that collaborators are actual people whose behavior is generally unpredictable. It becomes nearly impossible to construct a standardized setting for the assessment. Computers offer a sophisticated solution for controlling this variable: the computer agents act as collaborators in the assessment of real humans.

Scrutinizing the manageability of the data that is given by the assessment instrument unveiled the complexity of certain data. The previously listed suggestions for enhancements given by the experts would each bring a large increase in complexity to the instrument. Certain research approaches, such as the practice-based learning activities by Curcova et al. (2018), include the entire environment of a group to better understand intra- and inter-

group relationships in a collaborative problem-solving situation. The underlying learning-analytics system has specifically been created for collaborative scenarios and collects both ambient and live data within the environment. This confirms the experts' opinion that gathering complex data is a challenge, but manageable with a specific goal in mind.

Findings such as those published by Rosen and Tager, as well as of the OECD, were the initiating factors for this thesis: the belief that a human-agent approach can be implemented in a way that brings stability to the assessment and, even though natural language is a challenge, it is worth giving it a try. With PISA, a large-scale study has been conducted that gave meaningful findings; hence, it is obvious that further research can be based on these profound results in order to evolve the field.

### **3.4 Interview Summary**

In conclusion, the interview results emphasized the great potential of the field. Experts shared information and objective thoughts on the topics with extensive professional expertise.

The focus within given answers is mainly a reflection of the current state of the assessment and on best-case scenarios how the assessment could best unfold its potential. Above all, critical voices state the gap between realistic problem-solving scenarios and currently applied ColPS assessments that don't entail all parameters of a realistic scenario. The interviewees were encouraging and optimistic about the likelihood that integration of natural language to the assessment is a promising next step within this field. Constraints that come with this project are the maintenance of validity and neglect of automated scoring. In the current state, those constraints cannot be resolved due to the insufficient capabilities of the instruments when using semantic interpretation of the natural language. The importance of the inclusion of numerous factors has been emphasized, as has the high number of influences that are neglected in most of the current assessments. Within the interviews, several questions could be answered about important considerations applying to the field of ColPS. The empirical research conducted has influence on further decisions about the implementation of the artefact that is developed for this thesis. These assumptions can be read in Chapter 4.

## 4 Assumptions from Empirical Research

The synthesis of the research on the construct of collaborative problem-solving and conversational agents so far has brought various results to light. This chapter builds the bridge between previously conducted empirical research results and the prototype to be developed in the next step of this thesis, which represents the prescriptive study. Section 4.1 gives answers to research questions posed at the beginning of the thesis. Section 4.2 sets out the criteria that will form the scope of the prototype in order to describe the desired final state of the thesis.

### 4.1 Referencing the Research Questions

The literature analysis serves as a method for the research clarification (RC) which is part of the research design, in order to identify criteria of success for the prototype. Subsequently, the descriptive study presented considerations, given by experts, that have an impact on the construct of ColPS.

- *What is the role of agent-based systems for collaborative problem-solving?*

The agent-based system has a central role in the field of collaborative problem-solving. This is evident not only in the fact that research is being conducted as human-agent approach, but the interview results also show that agents are the most promising instrument with which to conduct reliable and valid research and assessment.

Especially for assessment scenarios, the agent serves as a static variable in order to obtain an instrument of high content validity. Artificial characters to be implemented in a scenario can be mostly neutral, or they may take on certain characteristics. The possibilities for influencing the instrument seem to be extensive; therefore, the agent-based system has a special role to play, particularly as an extensive variable.

- *What are the most common approaches and associated difficulties of assessing collaborative problem-solving?*

Most common approaches that are taken into consideration within this thesis are human-human and human-agent approaches. Equally, the literature and interviews confirmed that these two approaches serve best to assess collaborative problem-solving. While the human-agent approach serves as a valid and stable instrument to preserve predictable and controlled data, the human-human approach brings a wider variety in the results. It is more experimental but allows the observation of the natural construct of collaborative problem-solving. The human-agent approach exists in many different versions; thereby, many variables can be adjusted in line with the interests of the research, including the sort of interaction a test-taker performs, the straight-

forwardness of dialog, or the enhancement of the problem-space itself. The simulation of artificial characters attracts a surprisingly high level of interest within the field. To recap, these two approaches can be seen as antithetical constructs that both build an extensive and promising base to the field.

- *What influence does the implementation of a “natural collaboration” have on the assessments for collaborative problem-solving?*

An empirically valid answer to this question cannot be given at this stage of the thesis. Interview results let one assume that the increase of so-called “natural collaboration” can have a positive impact on motivation and participation of test-takers; however, interview results showed that the factor “natural” might endanger the internal validity and reliability of an assessment. Nevertheless, for certain research objectives, especially targeting the social-psychological aspects, it is the only instrument that can provide meaningful answers.

## 4.2 Criteria for the Development of an Agent

Results of the research clarification and descriptive study have been synthesized in the form of criteria, serving as a framework for the agent to be developed in the next chapter. These criteria represent desired enhancements that can be applied in a context of conversational agents and collaborative problem-solving. The following three criteria are a selection from many possible criteria that can be derived from the previous results of the thesis. Those that are selected promise to be the ones with the greatest impact on the construct. The factors of influence that emerged in the interviews thus serve as a constant, while selected properties of the constructs “conversational agent” and “collaborative problem-solving” serve as variables in order to produce valid criteria.

- **Allowance of natural language**

One argument that came to light within the interviews was the implementation of natural language, which can influence the student’s participation and which is a social skill of the collaborative problem-solving framework. According to Wooldridge’s definition (1995) of a conversational agent, an agent has to show reactivity; thus, the impact of natural language as a logical criterion in order to create a supportive artefact is a given (Figure 7).

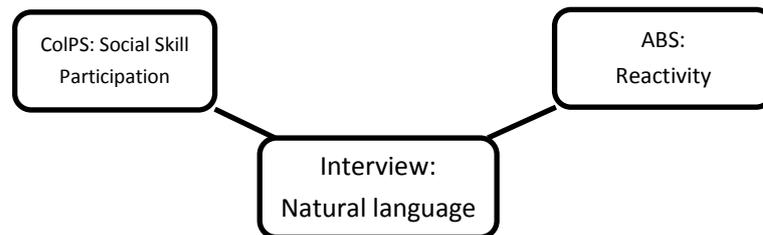


Figure 7: Criterion 1 - Implementation of natural language

- **Enrichment through interactivity**

Experts in the interviews emphasized the importance of interactivity, referring to the problem space in a ColPS assessment. Within the cognitive skills, the task regulation is one major criterion in the ColPS framework. Furthermore, IBM states in its definition of agent-based systems (Section 2.2) that an agent should “*employ some knowledge or representation of the user's goals*”. This forms a criterion that indicates that enriched interaction can contribute positively to an agent (Figure 8).

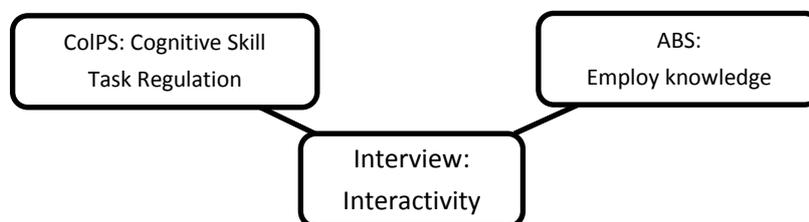


Figure 8: Criterion 2 - Implementation of interactivity

- **Enrichment through strong characteristics**

Thirdly, the interview unveiled the impact of characteristics that an agent can have. This certain impact can influence the outcome and perception of a collaborative problem-solving situation and might increase the difficulty of tasks. Social regulation is considered part of the social skills of ColPS activities. Within the definition of bots, characteristics can have an impact on the social ability that defines an agent (Wooldridge, 1995).

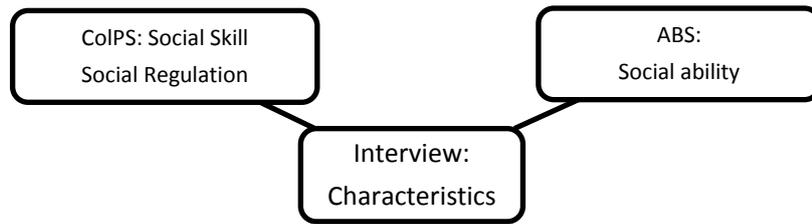


Figure 9: Criterion 3 - Implementation of strong characteristics

These three criteria build the base for the artefact to be created for this master's thesis. Their derivation from the clarified research context (Chapter 2) and the descriptive study (Chapter 3) builds a valid base for the creation of an artefact to be described in the next chapter.

## 5 Prototyping the Agent

This section represents the prescriptive study which describes the development of a working prototype that offers support to the research of ColPS. The prototype in this thesis is developed with Dialogflow, formerly known as API.AI, by Google. Its language processing is based on machine learning which allows the agent to understand a user's interactions as natural language and convert them into structured data. In Dialogflow terminology, the agent uses machine learning algorithms to match user requests to specific intents and uses entities to extract relevant data from them (Dialogflow, 2018). For this reason, the concept for developed dialogs is structured as conversation flows, a technique that is applied in many concepts of conversational agents. Further description of development decisions can be found in this chapter; an overview over the process can be seen in Section 5.1 followed by the presentation of the results in Section 5.2.

### 5.1 Process

This section describes the design process of the creation of the system. In order to develop a prototype that maintains a certain validity, the published Unit Xandar of PISA 2015 serves as the base concept. The development of the artefact follows the standardized ISO/IEC/IEEE 12207 “*Systems and software engineering – Software life cycle processes*” (Figure 10). The fulfillment of the entire process would exceed the scope of this thesis, but certain stages are followed in order to show a comprehensible approach to the development. The following describes the development process and task-based criteria of several steps in the form of a requirement catalogue.

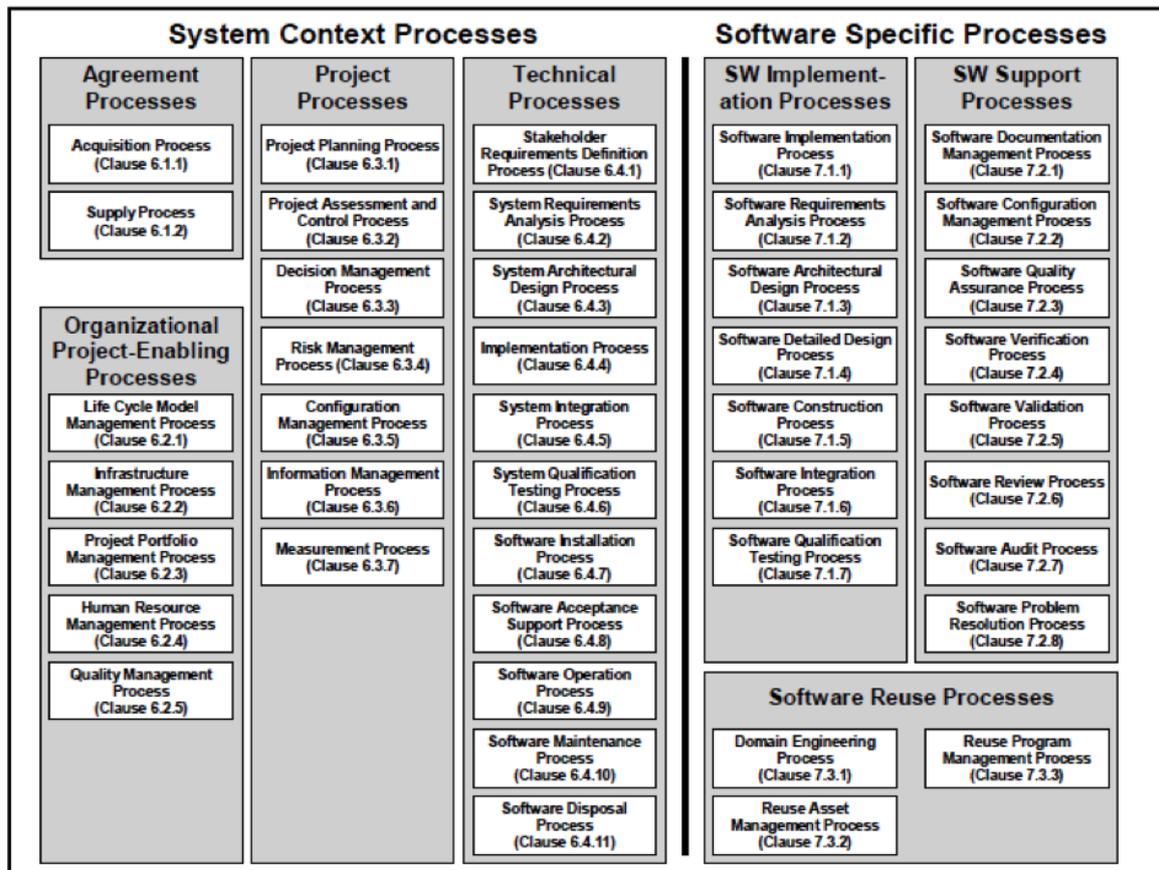


Figure 10: ISO/IEC 12207 "Systems and software engineering — Software life cycle processes" (IEEE/ISO, 2018)

To obtain the preset goals, the steps that had to be followed are: gather stakeholder requirements; proceed with a system requirement analysis; carry out system architectural design; perform implementation (implies software architectural design, software detailed design and software construction); and, finally, the system integration. The entire plan, conditions for the development and requirements can be seen in Appendix B.1: Software development plan. The following section (5.2) then describes the execution and results of this software development plan.

## 5.2 Execution

This section describes execution and results, created through following the software development plan (Appendix B.1).

### 5.2.1 Idea-Generation Stakeholder Requirements

#### Natural language

To obtain a solution that allows natural language, a conversational system has to be developed. The extent of natural language in a communication means that the user can talk in a natural manner to an interface with their own language. This can be done by providing a front-end to the user that is a chat client (Figure 11).

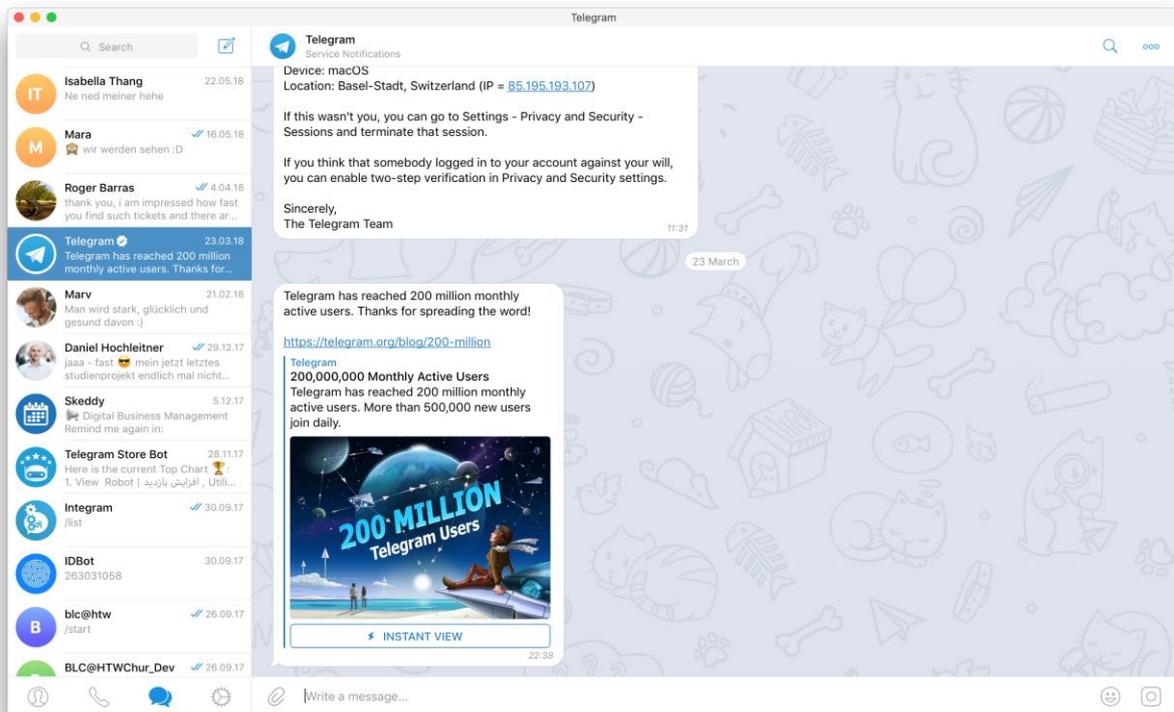


Figure 11: Chat Client Telegram (own representation)

## Interactivity

In the CoIPS Xandar Unit by PISA 2015, the system is divided into two parts. One part is the conversation space that contains pre-defined messages. The other part is the problem-space, where content is shown that provides the information of the task to solve. In Xandar, this is a scorecard with three buttons: Geography, People, Economy (Figure 12).

The screenshot shows the PISA 2015 Xandar Unit interface. At the top, it says 'PISA 2015' and 'Xandar - Introduction'. Below that, it says 'Part 1 - Directions'. The interface is divided into two main sections: 'Conversation Space' and 'Problem Space'.

**Conversation Space:** This section contains a chat window. At the top, it says 'Who's in the Chat' and lists 'YOU', 'Alice', and 'Zach'. The chat messages are:

- Alice:** Hi. I'm not sure about the best way to do this.
- Zach:** Let's just get going.

Below the chat messages, it says 'You are continuing the chat. Click on a choice below. Then click on Send.' The 'Conversation Space' also contains a 'Send' button at the bottom.

**Problem Space:** This section contains a 'Scorecard' table with three columns: 'Geography', 'People', and 'Economy'. Below the table, there are three buttons labeled 'Geography', 'People', and 'Economy'.

Figure 12: Interface of Xandar Unit in PISA 2015 marking the two main components of the interface: Conversation- and Problem Space (oecd.org, last access 01.08.2018)

Clicking the buttons activates the according test questions (Figure 11) per topic of the Unit.

**PISA 2015**

► Xandar - Introduction

▼ Part 3 - Directions

Your team has reached the following agreement.

- Geography** will be your subject.
- People** will be Alice's subject.
- Economy** will be Zach's subject.

Click on the symbols on the map to learn about Xandar and find the answers to the questions on the right.

When you find the answer to a question, click on the answer space next to the question and the answer will display.

When a question is answered correctly, a checkmark will be added to the scorecard.

To continue, click the button below.

[Click Here to Continue](#)

Scorecard		
Geography	People	Economy

Geography People Economy

What is Xandar's longest river?

What is Xandar's tallest mountain?

What is Xandar's rainy season?

What proportion of Xandar is desert?

Figure 13: Extended Problem-Space in the Xandar Unit of PISA 2015 assessment

The attribute of “interactivity” should at least fulfill the current interactivity of the unit at the same extent. A more graphical approach might increase the interactivity of an item. Furthermore, the conversation space should not be disabled during the collaborative problem-solving scenario. Leaving the test-taker the freedom to choose whether to gather information through conversation or through interactive components should increase the authenticity of the scenario.

### Strong Characteristics

The third focus is the implementation of strong characteristics into a client. The implementation of this criterion does not follow certain sociopsychological standards and merely provides a concept to demonstrate the impact a strong character can have in a problem-solving scenario. This concept hypothesizes the implementation of two contrary personalities that will demonstrate the phenomenon; thus, two personas have been developed that build the base for the personalities. Personas in this study are qualitative instruments that are defined by following the proposed approach by usability.gov<sup>3</sup> which can be seen as a state-of-the-art approach. Personas can be seen in the Appendix of this thesis (B.2).

<sup>3</sup> Usability.gov (<https://www.usability.gov/how-to-and-tools/methods/personas.html> / last access 18.07.2018)

## Conclusion

A system that fulfills the previously defined requirements had to be put into a conclusive context. To obtain a “big picture” of the expected result, a mockup of the system’s end-state has been created (Figure 14).

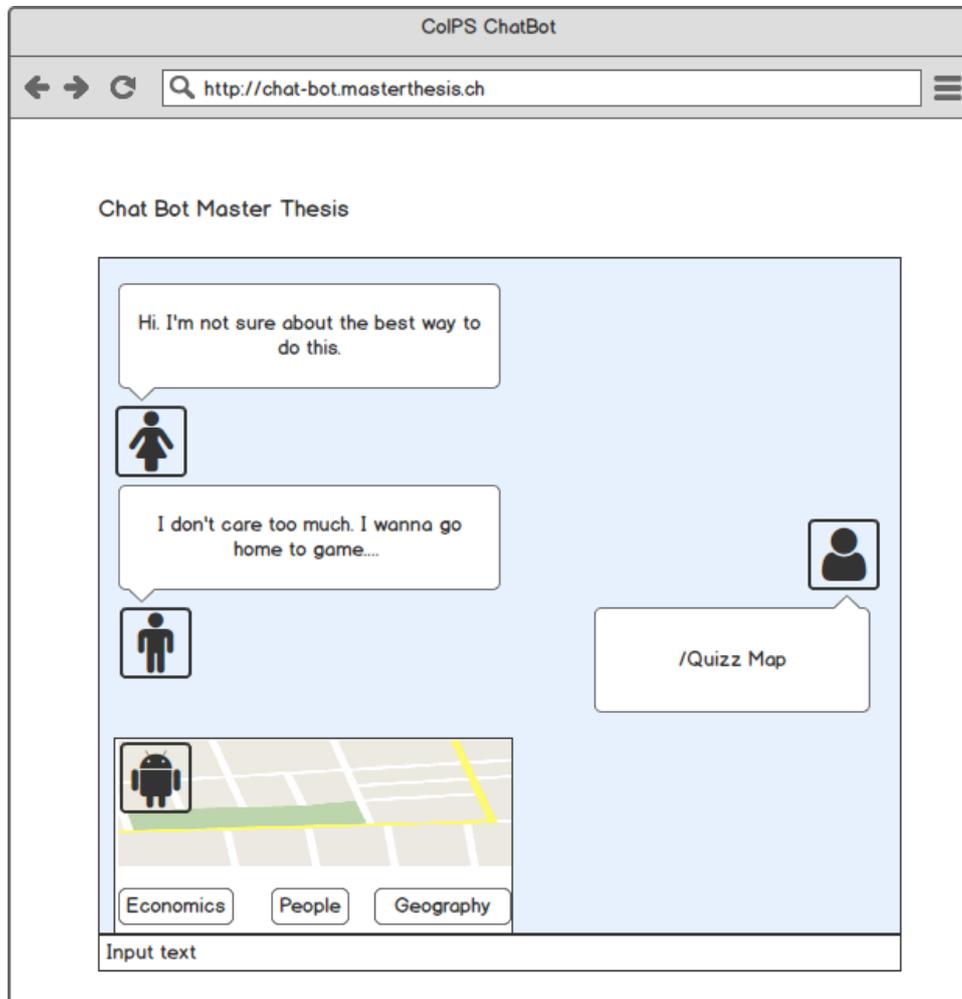


Figure 14: Mockup of the expected final state of the system

### 5.2.2 Matching System Requirements

#### Component 1: Conversation-Interpreting System

First, the conversation-interpreting system has to be determined. The market of so-called chatbot platforms is versatile. To gain an impression, and to find a system that could fulfill the previously set requirements, the overview of Davydova (2017) published on chatbots-journal.com was helpful. This article compared 25 platforms that offer the service to create a conversational system. Requirements of the platform were; free to use, support the English language, and integration of conversational clients that also offer free access.

With the given requirements, three of the twenty-five platforms suited the context of the purpose: Wit.ai, Kitt.ai, and Dialogflow. In the context of a research project, an open source platform would have been preferred due to data security reasons; hence, there currently seems to be no such platform that was able to meet all the requirements of this intended project. Due to time restrictions that is a given with a master's thesis, an extensive native development would have endangered the success of a project.

The platform of choice for this project is Dialogflow. It provides a well-developed environment with a comprehensible documentation to novices.

### **Component 2: Back-end**

The back-end system of this agent is inherited within the Google Cloud. This solution served as the most functional one in order to use the functionality of Dialogflow extensively.

### **Component 3: Conversational Interface**

Dialogflow allows the integration of the developed artefacts into many conversational systems such as Telegram, Facebook Messenger, Slack, Viber, Twitter and others. Conditions for the selection of a conversational interface include the possibility of implementing enriched content such as interactive cards that can trigger further actions of a system. Initially, Telegram was believed to be a suitable interface but has been discarded due to its limited access. A user can only access and use Telegram when signing up with a mobile number. To make sure the system to be developed can be accessible to a wider target group, this limitation excluded Telegram as suiting client; however, Slack meets all the criteria, and the creation of a workspace that can serve as a test environment for both developers and test-users that could then provide the base for further development.

### **5.2.3 Development System Architectural Design**

The prototype in this thesis has been developed with Dialogflow, formerly known as API.AI by Google. Its language processing is based on machine learning which allows the agent to understand a user's interactions as natural language and convert them into structured data.

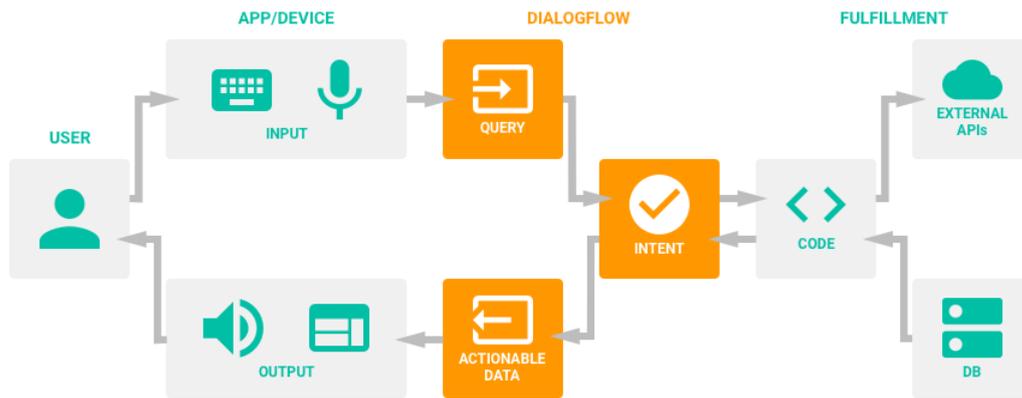


Figure 15: System components (Graphic by dialogflow.com, retrieved 03.08.2018)

## 5.2.4 Implementation

### 5.2.4.1 Software Architectural Design

The architectural design of the artefact to be developed can be explained as follows: Three components have to be implemented within a chat-based environment that offers the user a certain freedom and clarity of interaction. In this case, the two agents Anna and Kevin serve as agents for social interaction, and the “Xandar Quiz Instructor” represents the component to employ the problem space to the user (Figure 16).

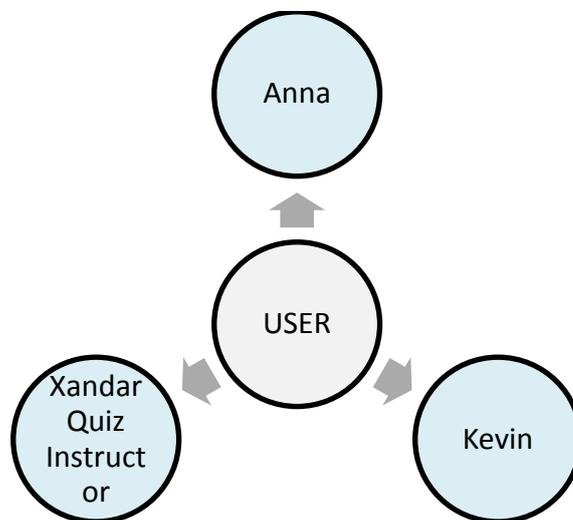


Figure 16: Architecture of the interactive component

### 5.2.4.2 Software Detailed Design

#### Conversation flows

In Dialogflow terminology, the agent uses machine learning algorithms to match user requests to specific intents and uses entities to extract relevant data from them (Dialogflow, 2018). For this reason, the concept for developed dialogs was structured as conversation

flows, a technique that is applied in many concepts of conversational agents. Dialogflow suggests the approach of devising conversation flows in order to have a profound concept of the goals of an agent. These conversation flows are presented in the following sections with an explanation of their context and connection to the original Xandar Unit. The logic that is forming these conversation flows can be understood as follows. A conversation has been formed with the preset conversation given by the PISA unit of Xandar. In order to maintain the impression of an actual conversation, the bots do not answer to text messages out of context. The suggested fallback intent by Dialogflow would be, e.g. "Can you say that again," which would imply that the interface is being used as a voice assistant. This is not the case, so the fallback intent has been excluded from this prototype. Furthermore, the idea behind the scoring was being maintained. When the user sends a message similar to the scored message in the original assessment, the conversation proceeds as planned. In order to simulate a more fluid conversation flow, a solution to the other case was necessary: What happens when the user is not saying something that could be directed to the right scoring? The avatars either give hints, hidden in the conversation, or the avatars give the scored answer themselves in order not to interrupt the flow of the task. As follows, this chapter briefly describes the implemented conversation flows. The graphical representation of the conversation flows can be found in Appendix B.3. The introduction how to set up the test environment, and the explanation of the task itself has been described on a separate introduction guideline (Appendix B.4) and on the webpage that is used as documentation to this thesis: <http://770695-2.web1.fh-htwchur.ch/demobot.html>.

### **Item 1, Part 0: Start Conversation**

In order to start the Unit, the user has to start the conversation. The initiation of the dialog is not part of the Xandar scoring guide, published by OECD in 2017. By greeting the other participants, the first explanation of the unit is given: the Xandar Quiz instructions can be triggered by typing "Xandar Quiz". The conversation flow to this section can be seen in Appendix B.3.1

### **Part 1, Item 1: Following Directions**

The first item in the Xandar Assessment is the introduction to the task. The user triggers the Xandar Quiz instructions and can see a card that describes the several elements to the text. Buttons that give further information about the tasks are "HowTo", "Geography", "People" and "Economy". Appendix B.3.2 shows the conversation flow to this section.

### **Part 1, Item 2: Understanding the Game**

After understanding the interactive Xandar Quiz instructions, the user should initiate further conversation. For example, the user could ask a question like "how can we do this?" The

conversation is being proceeded and ideally the user is suggesting making a strategy with the group. If the user is giving different answers, one of the agents will suggest the planning of a strategy in order to continue the conversation with the right logic. The conversation flow to this section is displayed in Appendix B.3.3.

### **Part 1, Item 3: Agreeing on a Strategy**

After the user and agents have agreed on the necessity of a strategy, they have to plan how a strategy could look that helps them in solving the quiz. The conversation continues with the same dialogs, given by the original Xandar Unit by PISA 2015. Again, the user can now give several suggestions, while the correct answer would be to ask the team members about a good way to find a strategy. Other messages will trigger the agent to suggest the definition of a strategy. The conversation flow as it has been described here can be seen in Appendix B.3.4.

### **Part 1, Item 4: Agreeing on a Strategy**

Team members were now in agreement that a plan for a good strategy is necessary. Again, the conversation flow is following the construct of the original Unit. The scored message is to suggest a division of the group, so that each one answers questions about one quiz topic. The suggestion of this solution would then lead to the next conversation flow. If the user is contributing differently, the agents will again give the suggestion of dividing the team to maintain the logic of the task. This is the last exemplary conversation flow that can be seen in Appendix B.3.5.

### **Reflecting on the creation of conversation flows**

The time available to this master's thesis restricted the implementation to five items of the original unit; however, it would certainly be interesting to have the complete unit implemented to this set-up. The completion thereby is a suggestion for further research, as well as the training of the environment to implement more scoring responses or back and forth conversation between the scoring events.

### 5.2.4.3 Software Construction

The conversation flows serve as concepts for the implementation into Dialogflow. The focus of the realization of the conversation flows was the creation of intents that can trigger the conversation. The interface for the creation of intents is shown in Figure 17:

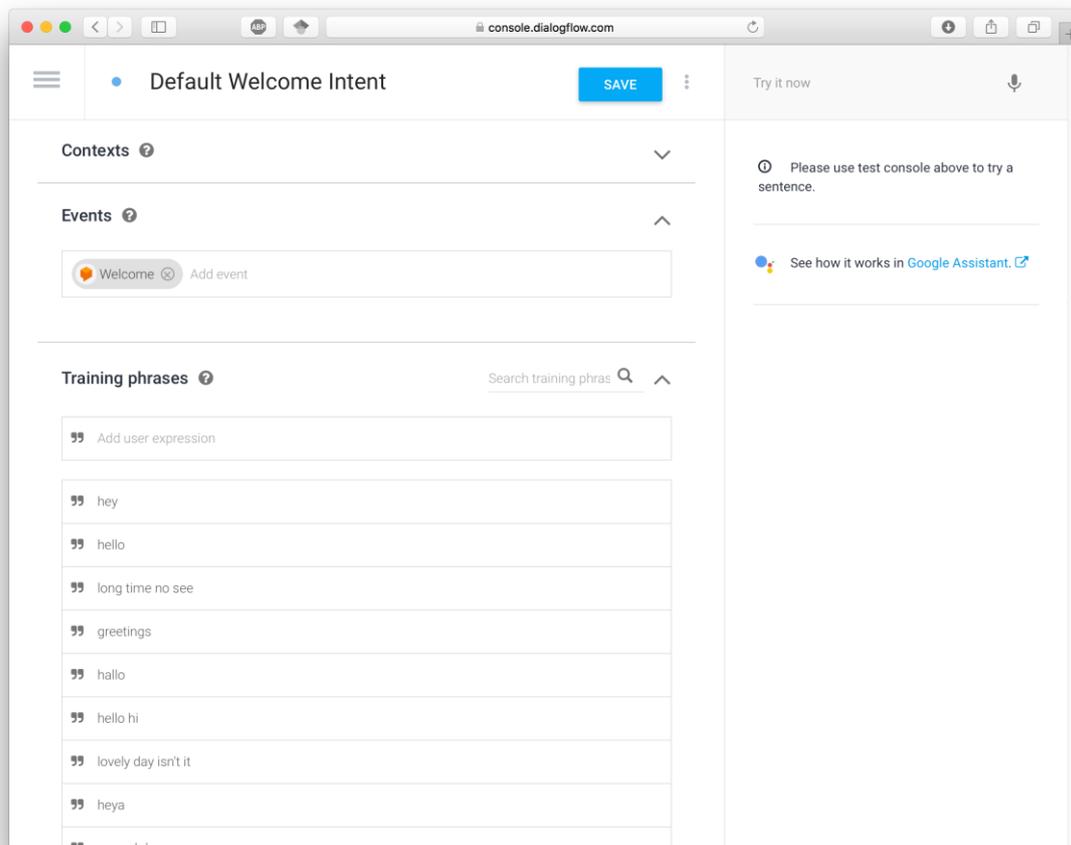


Figure 17: Intents in the Google Dialogflow system

An intent is a mapping between a message of a user and what action should be taken by the agent. Intents have the following dimensions:

- **Training Phrases:** Examples of messages, sent by a user. Can have parameters such as, for example, colors.
- **Actions:** Actions are sent to fulfillment once an intent is triggered (not implemented).
- **Fulfillment:** Code deployed through a web service to provide data to a user (not implemented).
- **Responses:** Responses that the agent gives to matching messages of a user.
- **Contexts:** Can be used to “remember” parameter values, so they can be passed between intents.

Each element of a conversation flow is implemented as an intent into the platform. To enrich the authenticity of a conversation, some intents were having responses with multiple messages. This subjectively resembles most the nature of a chat-based conversation.

Furthermore, the matching of the intents with messages was trained in several test-iterations (Figure 18).

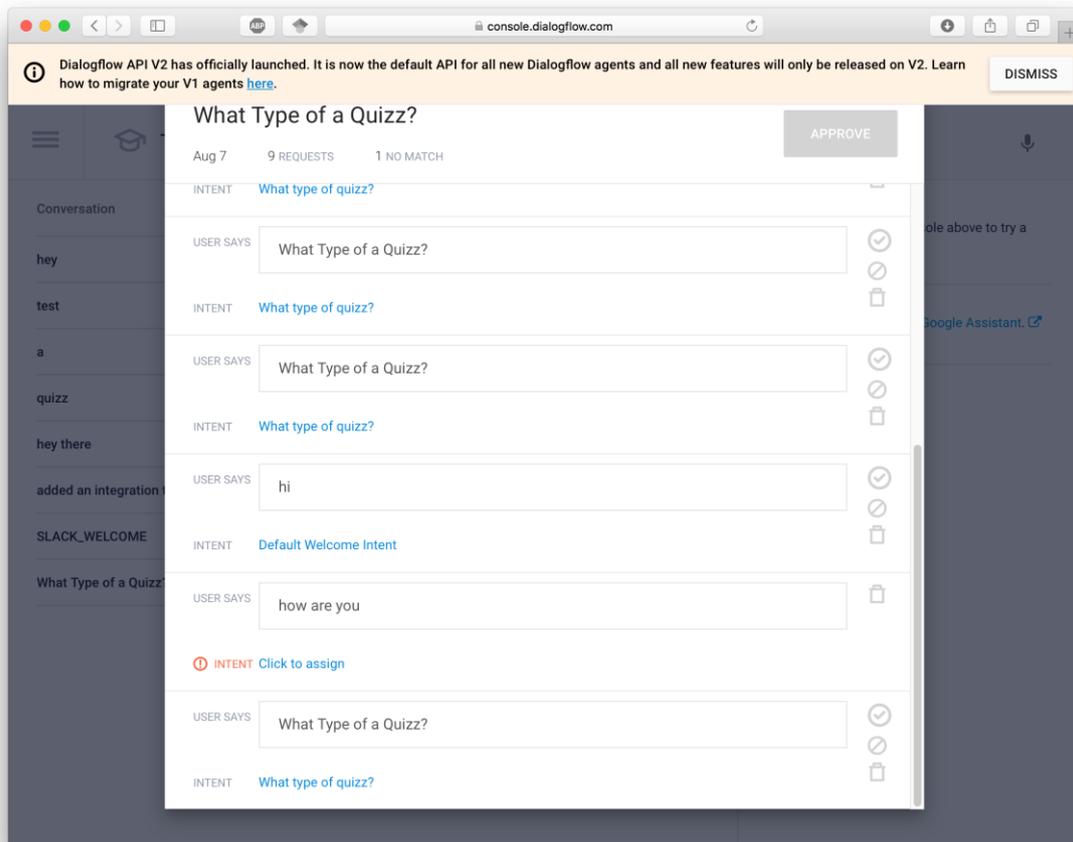


Figure 18: Training Interface of the Google Dialogflow System

### 5.3 Description of System Integration

The last step of this development is the integration of the software components into the system that reaches out to the end-user, namely the chat client. Slack offers a convenient creation of applications into their system by publishing an extensive API to developers. Consequently, three applications were created to allow the formation of a dialogue, accompanied by the interactive component that provides instructions of the quiz. These three applications can be installed independently into any Slack workspace. Adding them to a channel or to a group-chat is the final step to obtain a container for the simulation of the Xandar Unit. The final state of the environment to be created can be seen in Figure 19.

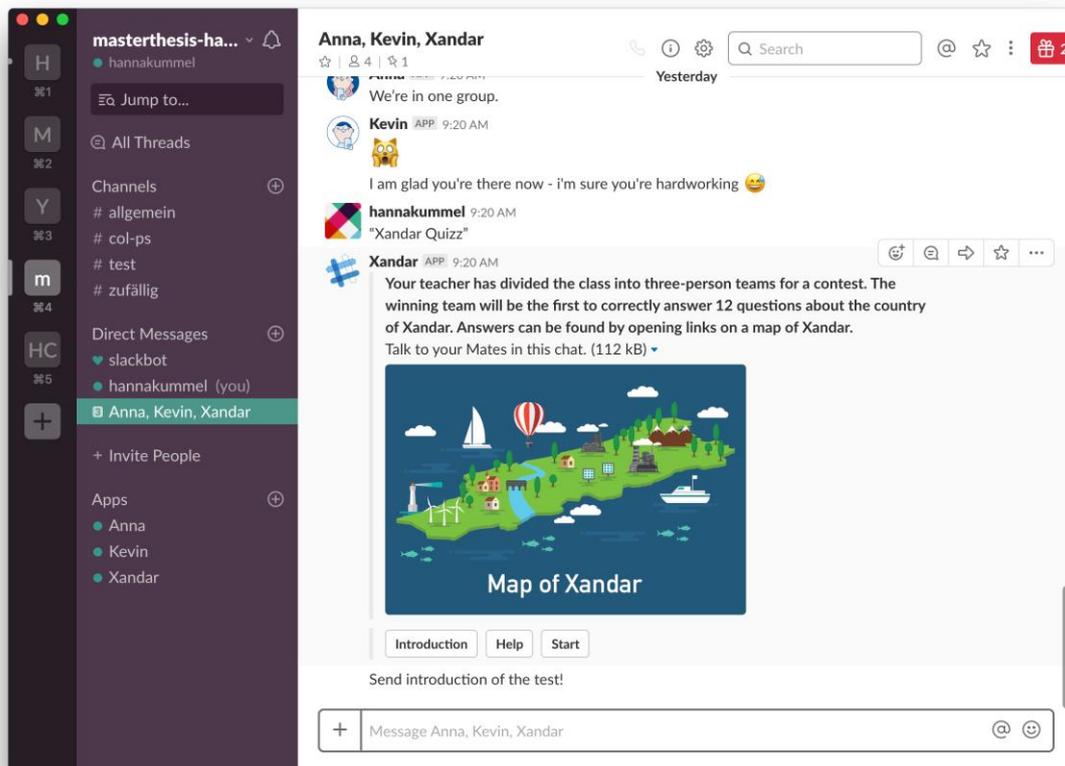


Figure 19: Xandar Unit simulation within a chat-based environment.

#### 5.4 Reflection of the Prototype's Creation

An evaluation of the retrieved system data through usage exceeds the scope of this thesis. Nevertheless, it is probably the most interesting part of the entire approach towards enrichment of collaborative problem-solving tasks through conversational agents. In order to estimate the appropriateness of this approach, the session data gathered through usage can be compared to predefined conversation flows (Section 5.2.4); yet, at this point, the agents would still need training in order to respond to a wider variety of conversations.

Although every single agent can be targeted in a conversation in a dialogue (which means not in a group chat with the other components), the private dialogue with an agent might not as yet be satisfying. The agents were created with a focus on functionality within the concepted test setup. The implementation of conversational structures is an elaborate process, and that is why it was necessary for this thesis to restrict the number of items to be implemented.

The preset criteria that formed the success of the implementation could be fulfilled; however, the weakest of all fulfillments is the implementation of interactive components. The initial idea of this thesis was to enrich the system by mini-games, but these could not be implemented because the implementation of such games is restricted to the Telegram client. Beyond the development that was planned, the implementation of emoticons into messages has been done.



## 6 Results of this Master's Thesis

The conducted research clarification served as a comprehensive instrument to this thesis. Subsequently, goals that had to be obtained in this thesis could be acquired. These were subdivided into three parts:

- (A) Aiming for the creation of a human-agent environment that follows Radinski's and Craswell's definition of a conversational system (2009).
- (B) Focusing on unveiling useful enhancements of a human-agent system.
- (C) Limiting the thesis to the development of five items of the Xandar Unit in order to showcase the achievement of the goals in (A) and (B).

Criteria resulting from the synthesis of research clarification (Chapter 2) and the descriptive study (Chapter 3) could be matched. A critical consideration of the achievement can be read within the reflection of the implementation (Section 5.4). At the beginning of this thesis, three research questions were introduced and two of them were discussed in depth in Chapter 4. The last research question to be answered in the introduction of this thesis is:

*Which influences does the implementation of a "natural collaboration" have on the assessments for collaborative problem-solving?*

This research question is not answered within this thesis, but several hypotheses that resulted from this research will be given as suggestions for further research:

### **Implementation of natural language**

H1.1: Use of natural language within a ColPS environment increases performance in participation of a user.

H1.2: Use of natural language within a ColPS environment positively influences the perception of reactivity of an agent.

### **Implementation of interactivity**

H2.1: Interactivity within a ColPS environment increases performance in task regulation of a user.

H2.2: Interactivity within a ColPS environment positively influences the perception of employed knowledge of an agent.

### **Implementation of Characteristics**

H3.1: Stronger characteristics of an agent within a ColPS environment increases the performance in social regulation of a user.

H3.2: Stronger characteristics of an agent within a ColPS environment positively influences the perception of social ability of an agent.

As stated in Section 4.2, the components that are likely to have an impact on the construct are, on the one hand, the dimensions of assessed CoIPS skills and, on the other, the dimensions that define a conversational agent. Thus, these two factors are presented as variables, while the constant of the instrument should be the criteria (natural language, characteristics, interactivity). This suggestion can serve as the last stage of the DRM framework that is underlying the thesis: a descriptive study that evaluates the artefact in the form of an empirical data analysis. For the valuation of this thesis, test-accounts are provided. The account data will be delivered separately with this thesis.

## 7 Discussion and Outlook

This thesis was introduced as a two-part research project that aimed firstly to examine the construct of ColPS as human-human and human-agent system and, secondly, to propose a solution that unifies advantages of both approaches of the interaction (human-human and human-agent). Both aims of this research project were met and the underlying methods were highly effective as a foundation for achieving the objectives.

Even though the Design Research Methodology is a rather new approach, it enables the development of artefacts with the support of valid and profound research results in a conclusive manner.

Presenting the argumentation of this thesis, there are several constraints that need to be mentioned. First of all, the literature review was a difficult task due to the limited population of publications in this field. Nevertheless, the research reports reviewed seemed to be of high quality. The empirical research that was conducted as expert interviews brought to light very interesting points of view about possibilities and limitations of the field. The number of interviews conducted could usefully be increased in order to obtain even more results and insights; nevertheless, both the research clarification and the descriptive study gave rich results to help answer the research questions that formed the scope of this thesis. The criteria in consequence, and set out in Chapter 4, served well to obtain a comprehensive artefact that matched the defined goals of this thesis. Following ISO 12207 norms when developing the artefact ensured the thoroughness of the implementation and the completeness of considerations. Thus, the setup of the prototyped environment seems to offer a promising construct as it can easily be taken forward, in both vertical and horizontal elaboration of this work. A vertical elaboration of this work could be the development of the full Xandar Unit and the reinforcement of preset criteria. The horizontal elaboration of this work could be the realization of further artificial characters with different attributes or the implementation of different ColPS Units.

Questions identified in this thesis that remain open are fundamental: the created artefact is currently implemented on a platform that does not ensure data security or any kind of privacy maintenance audit. The implementation into an open source framework that resolves this severe limitation of the artefact and allows it to be used for public research should be evaluated and would enhance the value of the entire research. Further topics that result from this thesis are emphasized in the reflection of the development of the artefact. Conducting a conclusive descriptive study through empirical analysis would complete the research of this thesis. Its evaluation can open further promising research directions to the field of collaborative problem-solving.



## 8 References

- Andrews, J. J., & Rapp, D. N. (2015). Benefits, costs, and challenges of collaboration for learning and memory. *Translational Issues in Psychological Science*, 1(2), (p. 182–191).
- Bergner, Y., Andrews, J. J., Zhu, M., & Gonzales, J. E. (2016). Agent-Based Modeling of Collaborative Problem Solving. *ETS Research Report Series*, 2016(2), 1–14. <https://doi.org/10.1002/ets2.12113>
- Blessing, L. T. M., & Chakrabarti, A. (2009). *DRM, a Design Research Methodology*. <https://doi.org/10.1007/978-1-84882-587-1>
- Bogner, A., & Menz, W. (2009). The theory-generating expert interview: epistemological interest, forms of knowledge, interaction. In *Interviewing experts*, (p. 43-80). Palgrave Macmillan, London.
- Care, E., & Griffin, P. (2014). An approach to assessment of collaborative problem solving. *Research and Practice in Technology Enhanced Learning*, 9(3), 367–388.
- Care, E., Griffin, P., & Wilson, M. (Eds.). (2018). *Assessment and teaching of 21st century skills: research and applications*. Springer.
- Camara, W. J. (2007). Standards for Educational and Psychological Testing: Influence in Assessment and Psychological Use. *Special Services in the Schools*, 2(2/3), 187–193. [https://doi.org/10.1300/J008v02n02\\_13](https://doi.org/10.1300/J008v02n02_13)
- Creswell. (2003). *Research design Qualitative quantitative and mixed methods approaches*. *Research Design Qualitative Quantitative and Mixed Methods Approaches*. <https://doi.org/10.3109/08941939.2012.723954>
- Cukurova, M., Avramides, K., Spikol, D., Luckin, R., & Mavrikis, M. (2016). An analysis framework for collaborative problem solving in practice-based learning activities. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, (p. 84–88). <https://doi.org/10.1145/2883851.2883900>
- Davydova, O. (2017, May 11). 25 Chatbot Platforms: A Comparative Table – Chatbots Journal. Retrieved July 18, 2018, from <https://chatbotsjournal.com/25-chatbot-platforms-a-comparative-table-aeefc932eaff>
- Derks, Daantje, Arjan ER Bos, and Jasper von Grumbkow. "Emoticons in computer-mediated communication: Social motives and social context." *CyberPsychology & Behavior* 11.1 (2008): 99-101.
- Dialogflow (2018). *Machine Learning | Dialogflow*. (n.d.). Retrieved August 3, 2018, from <https://dialogflow.com/docs/machine-learning>
- Dillenbourg, P., & Traum, D. (2006). Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *The Journal of the Learning Sciences*, 15(1), (p. 121–151).
- Fawcett, L. M., & Garton, A. F. (2005). The effect of peer collaboration on children's problem-solving ability. *British Journal of Educational Psychology*, 75(2), (p. 157–169).

- Franklin, S., & Graesser, A. (1996). Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. In *International Workshop on Agent Theories, Architectures, and Languages* (p. 21-35). Springer, Berlin, Heidelberg.
- Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. In *Behavior Research Methods, Instruments, and Computers* (Vol. 36, pp. 180–192).  
<https://doi.org/10.3758/BF03195563>
- Graesser, A., Kuo, B.-C., & Liao, C.-H. (2017). Complex Problem Solving in Assessments of Collaborative Problem Solving. *Journal of Intelligence*, 5(4), (p. 10).  
<https://doi.org/10.3390/jintelligence5020010>
- Greiff, S., Holt, D. V., & Funke, J. (2013). Perspectives on problem solving in educational assessment: Analytical, interactive, and collaborative problem solving. *Journal of Problem Solving*, 5(2), 71–91. <https://doi.org/10.7771/1932-6246.1153>
- Greiff, S., Wüstenberg, S., Holt, D. V., Goldhammer, F., & Funke, J. (2013). Computer-based assessment of Complex Problem Solving: Concept, implementation, and application. *Educational Technology Research and Development*, 61(3), (p. 407–421).  
<https://doi.org/10.1007/s11423-013-9301-x>
- Hao, J., & Service, E. T. (2015). Assessing Collaborative Problem Solving with Simulation Based Tasks Methods. *Proceedings of the 11th International Conference on Computer Supported Collaborative Learning*, (p. 544–547).
- Hikaru Komatsu & Jeremy Rappleye (2017). Did the shift to computer-based testing in PISA 2015 affect reading scores? A View from East Asia, *Compare: A Journal of Comparative and International Education*, 47:4, (p. 616-623), DOI: 10.1080/03057925.2017.1309864
- Kompella, R. (2018, February 09). Conversational AI chat-bot - Architecture overview – Towards Data Science. Retrieved June 11, 2018, from  
<https://towardsdatascience.com/architecture-overview-of-a-conversational-ai-chat-bot-4ef3dfefd52e>
- Krkovic, K., Greiff, S., Pásztor-Kovacs, A., & Gyöngyvér, M. (2014). New technologies in psychological assessment: The example of computer- based collaborative problem-solving assessment. *International Journal of E-Assessment*, 4(March), (p. 1–13).
- Lin, K.-Y., Yu, K.-C., Hsiao, H.-S., Chu, Y.-H., Chang, Y.-S., & Chien, Y.-H. (2015). Design of an assessment system for collaborative problem solving in STEM education. *Journal of Computers in Education*, 2(3), (p. 301–322). <https://doi.org/10.1007/s40692-015-0038-x>
- OECD. (n.d.). How does PISA measure students' ability to collaborate?
- OECD. (2017). PISA 2015 collaborative problem-solving framework. *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic, Financial Literacy and Collaborative Problem Solving*, (July), 131–188.  
<https://doi.org/http://dx.doi.org/10.1787/9789264281820-8-en>
- Pásztor-Kovács, A. (2016). Theoretical and Methodological Considerations in Assessing Collaborative Problem-Solving Competence: Results of a Pilot study. *Magyar Pedagógia*, 116(1), (p. 51–72). <https://doi.org/10.17670/MPed.2016.1.51>
- Ormrod, J. E. (1999). *Human learning* (3rd ed.). Prentice-Hall Inc.

- Mesmer-Magnus, J. R., & DeChurch, L. A. (2009). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology*, 94(2). (p. 535–546).
- Person, N. K., Graesser, A. C., Kreuz, R. J., Pomeroy, V., Person, N. K., Graesser, A. C., ... Human, S. (2007). Simulating Human Tutor Dialog Moves in AutoTutor To cite this version : Simulating Human Tutor Dialog Moves in AutoTutor.
- Polya, G. (1954). *Patterns of plausible inference*. Princeton, NJ, USA: Princeton University Press.
- Polyak, S. T., von Davier, A. A., & Peterschmidt, K. (2017). Computational Psychometrics for the Measurement of Collaborative Problem-solving Skills. *Frontiers in psychology*, 8, (p. 20-29).
- Radlinski, F., & Craswell, N. (2017, March). A theoretical framework for conversational search. In *Proceedings of the 2017 Conference on Conference Human Information Interaction and Retrieval*. ACM. (p. 117-126).
- Rölke, H. (2012). The ItemBuilder: A Graphical Authoring System for Complex Item Development. In *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2012* (pp. 344–353). Retrieved from <http://www.editlib.org/noaccess/41614>
- Rosen, Y. (2015). Computer-based Assessment of Collaborative Problem Solving: Exploring the Feasibility of Human-to-Agent Approach. *International Journal of Artificial Intelligence in Education*, 25(3), 380–406. <https://doi.org/10.1007/s40593-015-0042-3>
- Rosen, Y., & Mosharraf, M. (2014a). New methods in online assessment of collaborative problem solving and global competency. *International Association for Educational Assessment (IAEA) 2014 Conference*, (May).
- Rosen, Y., & Foltz, P. W. (2014b). Assessing collaborative problem solving through automated technologies. *Research & Practice in Technology Enhanced Learning*, 9 (p. 3).
- Rosen, Y., & Tager, M. (2013). Computer-based assessment of collaborative problem-solving skills: Human-to-agent versus human-to-human approach. *Research & Innovation Network*. Pearson Education.
- Schneider, P. (2014). Evaluierung der Chatbot-Beschreibungssprache „Artificial Intelligence Markup Language“, 31(April).
- Schumaker, R. P., Liu, Y., Ginsburg, M., and Chen, H., “Evaluating mass knowledge acquisition using the ALICE chatterbot: The AZ-ALICE dialog system,” *International Journal of Human-Computer Studies* ,2006, (p. 1132-1140).
- Selting, M & Auer, Peter & Barden, B & Bergmann, Jörg & Couper-Kuhlen, Elizabeth & Gunthner, S & Meier, Christoph & Quasthoff, U & Schlobinski, Peter & Uhmann, Susanne. (1998). A Transcription System for Conversation Analysis. *Linguistische Berichte*. 173. (p. 91-122).
- Stasser, G. & Stewart, D. (1992). Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of Personality and Social Psychology*, 63(3), (p. 426–434).

- vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A., ... Reimer, K. (2009). Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. 17th European Conference on Information Systems, 9, 2206–2217. <https://doi.org/10.1108/09600031211269721>
- von Davier, A. A., & Halpin, P. F. (2013). Collaborative Problem Solving and the Assessment of Cognitive Skills: Psychometric Considerations. ETS Research Report Series, 2013(2), i-36. <https://doi.org/10.1002/j.2333-8504.2013.tb02348.x>
- Wang, Y., & Chiew, V. (2008). On the cognitive process of human problem solving, Cognitive Systems Research. doi:10.1016/j.cogsys.2008.08.003
- Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. The knowledge engineering review, 10(2), (p. 115-152).
- Zehner, F., Weis, M., Vogel, F., & Spitzenstaaten, D. (2017). Kollaboratives Problemlösen in PISA 2015, 1–7. <https://doi.org/10.14459/2017md14017>

## 9 Appendix A – Interview Data

The appendix A contains three main elements to the thesis: interview guidelines (A.1), interview transcripts (A.2) and the matrix for interview evaluation (A.3).

---

### Appendix A.1 – Interview Guidelines

---

#### **Which problems do we have to face, assessing Collaborative Problem Solving in an Agent Based Environment?**

Interview guidelines for expert interviews

##### **Introduction: Thesis and overview**

Thank you very much for offering your time for this interview. First of all, I would like to give you a brief introduction to my master thesis, to this interview and its analysis. I'll introduce briefly the topics we will talk about, then we are ready to start the interview.

##### **About the Project**

This expert interview is part of my master thesis that I am writing at HTW Chur, aiming to earn the degree Master of Science in Business Administration Major Information- and Data-management. The expert interviews will be the base of the qualitative research of the thesis and help me to narrow down my research objective. In my thesis I aim to find an approach to a refined CPS assessment, conducted in an agent-based environment.

##### **Analysis of the Interviews**

If you agree, I would like to record this interview. Your answers will be handled as confidential information and will be analyzed anonymously. In case cites will be used in the analysis, all concrete information that might externalize a person's identity or affiliation will be removed.

**I appreciate not only your knowledge. Also, your valuation and opinion are a matter of interest in this Interview.**

Conducting this expert interview with you means that I appreciate your knowledge and skills in the research field. There might be questions, where you cannot provide answers with accurate, backed knowledge. Nevertheless, I would appreciate your answer, giving a subjective estimation or opinion on the question.

## **The interview process**

Questions in this interview will be subdivided into 5 topics.

TOPIC 1 – General Questions about CPS Assessments

TOPIC 2 – Matters of understanding and creating artificial characters

TOPIC 3 – Matters of understanding and structuring communications

TOPIC 4 – Matters of understanding the assessment instrument

TOPIC 5 – Sociodemographic Questions

### ***TOPIC 1 – General Questions about ColPS Assessments***

1. Do you feel confident with the current state of ColPS Assessment with agent-based systems (e.g. such as been used in PISA)?
2. Do you feel confident with the current state of technology, available to conduct collaborative problem solving with students?
3. Other than CBA Item Builder, do you know, or have you worked with other tools to create items for collaborative problem solving? If yes, what did you like or dislike?
4. If you could name the most important aspect of the assessment of collaborative problem solving, that needs improvement - what would it be?

### ***TOPIC 2 – Matters of understanding and creating artificial characters***

1. Can you briefly describe the interaction of a student with an agent in a current ColPS assessment?
2. How would you ideally expect an interaction with an artificial character in an agent-based system?
3. If you have to build ColPS Items now, is there anything else missing for the best possible implementation of artificial characters for ColPS Items?
4. How would you describe the perception of the student through the system? Does the student's behavior (anger, mistrust, fear or doubt) influence the artificial characters? Is there a channel to recognize the student's reaction (especially when messages are pre-defined)?
5. How does the artificial character influence the quality of the item or are there other aspects to be considered?
6. Do you think a personalization (as in personification) of the artificial characters help to enrich the item type?
  - a. visual personification (e.g. providing faces with mimics)
  - b. linguistic personification (e.g. talking in a rather informal language)
  - c. characteristic personification (e.g. strong characteristics)

**Topic 3 – Matters of understanding and structuring communications**

1. Can you briefly describe the communication, taking part in a typical ColPS assessment?
2. How would you ideally expect communication to happen in an agent-based system?
3. If you have to build a ColPS item now, is there anything missing for the best possible solution for communication within a CPS assessment?

**Topic 4 – Matters of understanding the assessment instrument**

1. Is there something missing when you think about currently used agent-based system, that a real life-interaction has, but the current assessment instruments do not have? If yes, what is it?
2. Did you have to process the obtained data that a ColPS (e.g.CBA Item Builder) Item generates (as in quantitative research) and how practicable was it for you to handle the data in order to conduct quantitative research?
3. Where do you see agent-based systems in your daily life? If you have seen some or are actually using them, which ones are you using? Which ones do you like the best and what do you like about them?
4. Researchers often speak of poor validity of the results. What exactly is endangering validity of CPS when being assessed digitally?

**Topic 5 – Sociodemographic Questions**

*Some questions about the interview partner and his research environment*

1. Where are you working now and what is your current position?
2. What is the main focus of your research within the ColPS field?
3. How large is the team you are working with?
4. Do you plan to work on the improvement of ColPS Assessment in the future?

**Final questions & interview end**

1. We reached the end of the interview. Did i miss a topic that we should have talked about or that you want to mention
2. Is there a closing word you would like to say?

Thank you very much for your time.

## Appendix A.2 – Interview Transcript 1

### A | Introduction

---

*Left out in recording due to connectivity problems.*

### B | Main Part Interview

---

#### Question 1.1

Interviewer: Hello?

IP01: Yeah, hello? [crosstalk 00:00:38]. Okay, okay. **So, when it comes to the question would I feel confident with the current assessment of Collaborative Problem-solving, in particular in PISA 2015, I would say to some extent.**

IP01: So from an assessment and a problem-solving perspective I would say yes, but I do know there is a lot of research that is about collaboration and more in terms of social psychology, in terms of truth dynamics, collaborative learning, education, and criticism from this field, or from those fields would probably mean that they would say that PISA 2015 is not natural enough. It does not allow for a wide enough range of communication.

IP01: Which I think is a general criticism when it comes to agents but then again the PISA assessment was very restrictive in terms of also what the agents were able to do.

IP01: Given that something like this has never been assessed in a large scale assessment, at least not in PISA and not on such an international scale, I think it was a first good shot. I think also given the constraints in terms of you need a standardized assessment, not allow for a too wide kind of field of collaboration or cooperation within the tasks, and you need to be able to score the individual performance and not the team performance.

IP01: I think what was done in PISA 2015 was a good first shot. At the same time many improvements could be made, but also when you introduce improvements in terms of you kind of prod of the assessment space and you make it more natural or you allow for more types of collaboration.

Interviewer: Yeah.

IP01: Always kind of get other problems at the same time.

Interviewer: Yeah. I was reading your publication about these problems you are addressing, and I ... Yeah, I feel what you are talking about is something I can understand very well. Because you were talking about many problems that you have to face already, during the assessment, and that it is not easy to provide solutions to the current problems that there are existing.

---

**Question 1.2**

- Interviewer: So for my field, or for me personally, I would like to ... Since I'm not a social psychologist, I'm not a pedagogue or nothing like this, so I would like to focus a little bit more on technology. **So, if you consider current state of technology that you have available to conduct Collaborative Problem-solving with the students, how do you feel about the current state?**
- IP01: Well, given that I am a psychologist, probably my view on the technology is somewhat limited, in terms of what I know about the current state of VR. I think that it's always a bit different of what can be applied on a broad scale and what is technically possible.
- IP01: I would probably think that much is technically possible and that you can have very intelligent agents that interact in an almost natural way with students, but that might not be at the stage where you can kind of implement it into a study such as PISA.
- Interviewer: Yeah, okay.
- IP01: I know that the Collaborative Problem-solving extra groups, so the OECD, has always scientific extra groups.
- Interviewer: Yeah.
- IP01: And they were basically different views. Some guys that had more computer science background, they basically said, we can't have an agent that also gives the impression of being almost a natural type of communication. While others said, no, nothing can basically ... Well, nothing can be as natural as just kind of working together with another student.
- Interviewer: Yeah.
- IP01: So I guess personally I would be somewhat reluctant in terms of, well, in terms of the claim that we can really mimic natural collaboration with agents.
- Interviewer: Yeah, besides natural language, probably there are other approaches to, yeah, to get towards a more natural conversation, and that's something that I would like to figure out a little bit more.
- 

**Question 1.3**

- Interviewer: Did you ... But that's something later on. You were working, probably with a CBA item builder, have you personally worked with the item builder to create items? Do you maybe know something or some other tools that are existing in order to solve the same problem?
- IP01: Well we've worked a lot with the CBA item builder but not for Collaborative Problem-solving. That was actually ... So we never actually offered any items, but it was done by the test developer, by the educational test developer.

- IP01: They basically do item by item implementation into an overarching framework, but they don't have this kind of offering tool that the CBA item builder is supposed to be.
- IP01: But I started working with the item builder maybe 10 years ago, and I have grown critical, to some extent, as I see that it is very difficult to actually come up with offering tools, the more specific or more demanding an item type gets.
- IP01: So I think like in the case of Collaborative Problem-solving for PISA 2015 it was a good approach to basically implement it item by item, because the demands became at some point so specific that it's quite difficult having a generic platform that allows a non-computer scientist or a non-program, or a non-IT person to actually work with the operating tool.
- IP01: It was also with the CBA item builder at a point where we wanted student assistance to work with the item builder, and they were able to do that, but it took a lot of time to really acquaint themselves with the item builder.
- Interviewer: Okay.
- IP01: Once they were acquainted enough they usually left, so that was a bit of a difficult situation.
- 

#### **Question 1.4**

- Interviewer: Okay, I understand. Okay. If there would be one most important aspect of the assessment of collaborative problem-solving, that needs improvement, what would it be? Would it be rather the technical aspects? Or do you have other aspects in mind that you feel are the most lacking ones?
- IP01: Well I mean the technical one is an obvious one. Another one would be to reduce reading load. Because when we talk about PISA 2015 it's a high reading load on all the units, but it's not supposed to be reading related. It's like being able to read is not core of the construct.
- Interviewer: Okay.
- IP01: I think somehow widening the collaboration space from a content perspective, in terms of making it not so reading dependent, allowing for little type of different collaborations, that would be important.
- Interviewer: Okay, if you're talking about different collaboration, what do you have in mind?
- IP01: Can you say that again?
- Interviewer: When you talk about sort of a different collaboration, what do you have in mind? Except for the agent based approach, for example, or do you still mean the agent based approach but different in some other kind of way?

IP01: Yes, I do mean the agent based approach. I mean in PISA 2015 it's just so limited in terms of, it's multiple choice, it's all region.

Interviewer: Yeah.

IP01: It's a very, very limited type of interaction also, number of steps is very limited. I think the challenge would be to widen this in terms of, you know, when you look at tutoring systems or something like that, you know, you have some kind of different interaction there and kind of widening that without losing the assessment reader that is needed.

---

### Question 2.1

Interviewer: Okay, I understand. Okay. Good. To get to the next topic, I would like to understand a little bit better how the creation of the artificial characters could be ideally? I'm more referring to the current state, how it has been, having two artificial characters, or actually I'm not even addressing the amount of artificial characters, but in the end, I would like to ask you about your findings in problem-solving itself, where you were having an approach with three different types of interaction.

Interviewer: But that's for later on. For now, **I would like to ask you to briefly, maybe just very briefly describe the interaction of a student with the Collaborative Problem-solving assessment.**

IP01: In 2015? PISA 2015?

Interviewer: Yes. Yes.

IP01: As I mentioned already, it's quite limited. It's all check based, and it's multiple choice based.

Interviewer: Yeah.

IP01: The student gets different answer options, and then you know, something happens in between, like the agents say something, or they do something. Yeah, so it's multiple choice, it's check based, and usually it's kind of not ... It's quite a few steps until a task is solved, so it's not like a longer flow of interaction.

Interviewer: I understand. You said if I'm referring to 2015, have you seen different approaches?

IP01: Well I know that, as I mentioned, the PISA collaborative learning of social psychology, I mean they have very different approaches, I think most of them are actually interactions with actual people, but I think some of them also use some kind of intelligent environment. Then there was of course the ATC21S initiative.

Interviewer: Yeah.

IP01: Which was also a large scale initiative which used real collaboration, but that has been also suffered from quite a few shortcomings that, as far as I know have not been solved.

- IP01: In a way it's like we need to find a good balance between making collaboration natural, which then makes it very uncontrolled, and making it at the same time, making it sufficiently standardized to be able to draw an assessment out of that.
- IP01: It's in a way a juxtaposition, and so ATC21S went very far on the side of making it natural, and less practical, and PISA went very much on the side of making it standardized.
- Interviewer: Yeah.
- IP01: I would think maybe some kind of just a bit more full would probably be a good next step.
- Interviewer: Okay. That's very good. I also think that yeah, maintaining the validity of the results is something that's very important and I see that there is lots of writing in the field that this is a huge issue, which makes it that hard to get more into a ... To make a step towards natural environments, or more natural situations. Yeah, I think that's a very hard step to get towards to.
- 
- Question 2.2** see above
- Question 2.3** see above
- Question 2.4**
- Interviewer: How do you feel the perception of the student himself, through the system? Is there anything existing that the system recognizes and reacts to the student behavior? And do you think it is necessary?
- IP01: Again, in 2015, obviously the agents react specifically to the choice the student made, but of course it's a discreet number of choices.
- Interviewer: Yeah.
- IP01: Sometimes basically the student, or well the agent reacts irrespective to what the student chose, sometimes it's targeted at that.
- Interviewer: Yeah?
- IP01: I think that is necessary, I mean kind of the natural flow of communication is that, you know, there is an input and there is an output, a response that we have to reinforce. That's why that was also so constrained in PISA 2015 because I understand you need highly developed agents in order to enact like open input and then a targeted response, and this was always imperfect.
- Interviewer: From my side I was thinking about, for example, if you allow natural language, there are already good technologies to structure language, and you have, for example, approaches like IBM has offers tone analyzers that actually offer an interpretation, a quite solid interpretation of language and it's meaning.
- Interviewer: So I think there are interesting technologies available, but of course it's a question of how well they work and how good they can be

implemented in order to maintain the validity of the item type. But I think there are interesting approaches. I don't know if you heard about IBM solutions on language interpretation?

IP01: No, but I am aware that there are these type of approaches. It was not feasible for PISA just because of the number of languages.

Interviewer: Yeah, that's true. Yeah, I read it.

IP01: Yeah, but having some kind of study like this, in a specific language context probably would make a lot of sense.

---

### Question 2.5

Interviewer: Yeah. Okay. Good. Do you think that a personalization of the artificial characters would help to enrich the item type? Such as a visual personification, or the linguistic that we already discuss now quite a lot, or in the end a characteristic personification? I think for the linguistic personification it's clear that's something that we were having talked about now a lot, but for example, having a visual personification of the artificial character, do you think this is an approach that might enrich the item?

IP01: I think so. I mean I think that when you look at these tutoring things, usually they have some kind of face. I mean it's always like they might even be some kind of animal or whatever, or some kind of tool. I would think that is, yes. I also think this can increase the engagement of the student. That's something that as well was missing in the PISA 2015, I think if you go through the items it's very ... You know, they're not very rich, graphically.

IP01: On the other hand, of course, you know, with these visual aspects you also introduce some kind of noise as well. But having some more or less neutral kind of visual face, that where you would think that people wouldn't react too strongly to, I think that would make sense.

IP01: As for the characteristic personification, I would think that, you know, you would need to define what you think a good collaboration is. That would probably involve being able to deal with different characteristics, so I would see this more as a way of content validity of the assessment, to say well, we theoretically decide that somebody who's good at collective problem-solving, you know, can deal with the strong personalities or with the unmotivated personalities, and have, for the sake of content validity, these aspects included.

Interviewer: Okay. Yeah, that's true. Okay, thank you so far.

---

**Question 3.1** See above

### Question 3.2

Interviewer: Which also belongs to the artificial characters, but it's rather own topic, is the communications that do take place. Yeah, I can ask you again about the communication, how it is in a CPS assessment, but you

already said, it's very structured, it's very static, it's very ... It's only multiple choice, so the communication is actually not a natural communication, but besides opening to natural communication from the side of the bot and from the side of the student, how would you, ideally, expect the communication to happen?

IP01: Well that is a good question. Because of course everybody could easily say having it as open, as natural as possible. I mean that is of course that I can say this, and I'm going to say that's wishful thinking.

IP01: Because that's not going to happen, and particular not going to happen in assessment context. So what I would probably like to have is at least some more options. As I mentioned, they should be less reading dependent, or less written in a way. So there should be the possibility to also communicate through some actions, visual actions. I know this is to some extent implemented into the ATC21S task, that students just kind of work together on a task as well, but they had open chat, which made them run into immense problems.

IP01: So, it's kind of a question, how can you make interface that doesn't take long for students to understand? That allows a bit more action than just four multiple choice answers, but doesn't allow for open chat, because this is too complicated. I don't have a perfect solution for that, because if I did I would already kind of develop items.

Interviewer: Yeah.

IP01: In this direction, but I think it's kind of balancing these different aspects of better station, not too reading depending, giving space for collaboration.

---

### Question 3.3

Interviewer: Yeah, okay. That's almost answering, but maybe I'm still asking you, if you have to build the collaborative problem-solving item now, and not 2015, is there anything missing for the best possible solution for the communication? What is missing for you as a tool, probably? What do you think you would miss?

IP01: Well, I mean I would have probably said that in 2015 as well, because just the time for the item development in these [inaudible 00:22:32] assessment is so short, so you don't really ... You need to come up with something within weeks, basically.

Interviewer: Yeah.

IP01: So I probably would have ... Actually, I think some of the ATC21S tasks, they kind of ... You know, the students have different information and then they kind of work together on something, so it's a bit more of interactive collaboration, because it's very kind of step by step, it's very discrete.

Interviewer: Yeah.

- IP01: I think if I wanted one thing to be implemented then it's a bit more continuous flow of interaction.
- Interviewer: Okay. If you talk, say interaction, what do you actually have in mind thinking about an interaction? Is it for example rather interaction as I am drawing as well? Or do you have something special in mind?
- IP01: No. I mean I probably wouldn't say it's only kind of talking, but it's also doing something, on some kind of search environment. I mean if you look at I guess classical gaming, you will find many examples of where people have to work together online, whatever, so some of this.
- Interviewer: So maybe doing research on how people in reality work together as in, they sketch, they write, they create, yeah, idea maps, or something like that, to implement that would be helpful to create better communication and not only language based communication, if I understand you right?
- IP01: Absolutely. Absolutely.
- Interviewer: Okay, that's very nice. I also think that would be ... Yeah, a very good approach to actually try to structure natural groups or natural problem-solving and maybe enhance it with some more tools to get to a result. It sounds very good. Yeah.
- 

#### Question 4.1

- Interviewer: The last big topic is for me to understand the assessment instrument itself. But I already started a little bit in the very beginning. It all gets together here. If there is something missing on the agent based system that a real life interaction has, that is something where we just came to. Do you have other ideas besides what we were just mentioning? For example, adapting from real life situations?
- IP01: Well, no. I mean I think what I find important is to kind of make a good distinction between a theoretical framework on collaborative problem-solving, which would probably kind of encompass all situation that you could think of, which we collaborate as human, and an assessment framework, which will always be limited. No matter how rich you kind of develop the environment.
- Interviewer: Yeah.
- IP01: You know, again, I could talk in terms of wishful thinking, we need this, we need that, but given realistically what you could have, even with a lot of time for development and a lot of time for assessment, you will always be limited. I think that's not a problem, per se, it just needs to be spelled out. There's specifically what these limitations are.
- IP01: Of course the PISA framework has been very restrictive in terms of the type of interaction it allows and basically this could be theoretically why [inaudible 00:26:24] definitely, but I think we discussed our suggestions on where I would widen it if I were to say, what would be the first aspect where it could be brought or opened up.
-

**Question 2.2**

- Interviewer: Okay. Okay. Now I'm shifting over to a totally different situation, so, maybe I would like to see if you ... Have you had to work with the actual data that the CBA item builder generates after the assessment? Have you had to handle the data that was generated through the assessment? And how practicable was it for you to handle the data that you got in the end in order to conduct quantitative, or maybe also qualitative research?
- IP01: Well, so with the results of collaborative problem-solving I'm not even aware which items are run in the item builder. I am not aware of any item on collaborative problem-solving in the item builder. We've not worked with those.
- IP01: We did some studies with the original PISA items and also with additional you know, item by item implemented collaborative problem-solving items, and there we did work with the project data, I think a very general notion is that you easily get lost in processed data and you just need to know beforehand what you are looking for, or you need to be very good in educational data mining.
- Interviewer: Yeah.
- IP01: Even there my impression is that you still need to know what you're looking for.
- Interviewer: Yeah.
- IP01: We did work with other items in the item builder, with process data and that was also ... We had at the end of the day an extraction tool, but there again the [were 00:28:16] of kind of deciding and theoretically deriving what we want from the process data was kind of the most important part.
- Interviewer: Yeah.
- IP01: Yeah.
- 

**Question 2.3** --

**Question 2.4** (left out later on)

- Interviewer: Okay. Talking on this, I can actually skip the next questions, but if I see it right, you and other researchers, yeah, often speak of ... Yeah, also of the poor validity that you get from the results then. Is there something that you see ... Is it because of the not very structured approach of the data? Or not of the lacking ... Yeah, maybe I would say ... Can I assume that the structure of the data is maybe something that is not very transparent?
- IP01: Yes, definitely. Yes. Absolutely. I don't even know whether it's publicly available, and that is usually also not because of the OECD wants to hide anything, I hear this notion sometimes that people get, I guess, a

- little bit paranoid, that you know, it's not [inaudible 00:29:37] information, it's just lack of resources, you know?
- IP01: It's like you know, if you want to structure this data but you don't know who's going to work with it, so somebody has to pay for it, and it's a lot of work.
- IP01: We did an initial study with the 2012 process data for the OECD and then we realized that all processed data, so on the creative problem-solving, not collaborative problem-solving, creative problem-solving, PISA 2012, and there we realized how much work it really is to just simply prepare processed data.
- Interviewer: Yeah, yeah.
- IP01: Again, I also know from the ACT21S people, if you talk to them they would not agree with that, but I have never seen any data, any publication ... They have this open chat, and they try to analyze it, and the best way they could come up with for doing it was the doing it manually. You know? They were looking at the chat, the chat input from both students and they had coders, who were coding.
- Interviewer: Yeah.
- IP01: Which is incredibly resource ... Which is incredibly costly.
- Interviewer: Yeah, yeah, I think so too. Besides the data, what do you think could also, here I'm writing about the validity, but what else is endangering rather the outcome of the collaborative problem-solving when it is being assessed digitally?
- IP01: Can you specify on that?
- Interviewer: Okay. How, or ... Yeah, if you're assessing it in a digital way, the collaborative problem-solving ... No, the question doesn't really make sense.
- IP01: Topic four, the fourth question?
- Interviewer: Yeah, it is the fourth question. I would say what exactly is endangering the validity of the collaborative problem-solving, but it is not really ... I was ...
- IP01: I guess the ... I guess the answer to that would be, empirically the validity has never been shown.
- Interviewer: Yeah.
- IP01: There needs to be some evidence that actually people that collaborate better in these tasks are also in real life better collaborators.
- Interviewer: Okay.
- IP01: That has not really been shown, and I think this also ... I think this probably goes for many research items. So you either need a strong content rationality, you know, [inaudible 00:32:22] is collaborative

problem-solving, from a content perspective, or you need this empirical evidence. I think both is missing. In a way.

Interviewer: Okay, no that's interesting. Now I'm glad that I still asked it, because I was thinking, oh maybe this question is not really suiting very well anymore, but it was just perfect. Thank you very much.

---

### **Question 5.1 – 5.4 are not part of the transcript due to anonymization and data security**

---

.....

---

### **C | Final questions & Interview end**

---

Interviewer: That's it actually. Now I'm at the end of my interview, and of course in the end I would like to ask you if I forgot to mention something? Did I miss a topic that I should have talked about with you?

IP01: No. I think you captured most of the important stuff.

Interviewer: Okay.

IP01: I mean of course it's important, I think, who the other experts are. I would hope you have kind of a diversity of people, that would probably help your research. [crosstalk]

Interviewer: Is there anything else that you would like to say before I at least end the-recording?

IP01: No, all good. I wish you good luck with the master thesis.

Interviewer: Yeah, thank you very much. I'm getting more and more, let's say I'm grabbing hope to get somewhere.

IP01: Excellent.

Interviewer: And I'm very, very thankful that you were offering me your time. It was extremely helpful and were amazing opinions that you could give me. So thank you very much.

IP01: You're welcome.

---

## Appendix A.2 – Interview Transcript 2

### A | Introduction

---

- Interviewer: First of all, of course I want to thank you very much for your time, to conduct the interview, because you help me a lot by, with this. [crosstalk]
- IP02: I hopefully will.
- Interviewer: Oh yeah. You cannot do anything wrong. From your side, I'm the one who has to handle it then.
- IP02: Okay.
- Interviewer: To talk to you about the project, as you know, it's my master's thesis. I aim to have the Master of Science in business administration, afterwards, hopefully. The expert interview, they are part of the pre-study. The pre-study, I would like to figure out what are general problems, approaches about collaborative problem-solving. I have some brief ideas in mind where I can dig in in detail, when it comes to the main study of my master's thesis, but the pre-study is really there to have a qualitative research on the topic.
- Interviewer: If you agree, I would record this interview?
- IP02: Sure.
- Interviewer: That's very nice, thank you. I will send you the transcript afterwards.
- IP02: Very nice.
- Interviewer: For you, important to know is that I will anonymize everything that we are saying now, and I will really break it down to the point, what we are discussing, and even in my analysis afterwards, there will be no traceability about the information.
- IP02: Yeah. Actually, I'm not worry about that, because I guess anything I would say to you, will soon be published, or at least this is our hope. Anyway, all of my ideas, I'm will to say out loud. I will say, hopefully, yes, in an article. Which, we will just about to submit.
- Interviewer: Okay. That's great. [crosstalk]. Okay. As I said, I will send it to you anyways afterwards, so you can see the transcript.
- IP02: Sure. Okay.
- Interviewer: It just takes a while, I guess about a week or something until I can send you the transcript.
- IP02: Okay.
- Interviewer: Yeah, in the end, of course, is very important to say that I appreciate your knowledge, it's very valuable for me. If there is anything you cannot answer on, it's not a problem at all. I just appreciate estimation or an opinion that you have on the question. You can just say, this is

not topic I was working on, but I would say, just to give me a brief idea about your opinion on it. That also helps a lot.

IP02: All right.

Interviewer: Okay. Then that's already it. The interview topics that I like to talk about are, first of all, general questions about the assessment, then matters of understanding the artificial characters that you are currently or have been in PISA 2015, as in the agents, because they have, as I've seen, there is not really an artificial character, but rather that agents that were implemented. I'm going with the terms as a artificial character. Then, matters of understanding how communication has been constructed in the assessment. Then, in the end, understanding the assessment instrument. Last but not least, some questions about you and your team.

IP02: All right.

Interviewer: Okay. Then, let's get started.

IP02: All right.

---

## **B | Main Part Interview**

---

### **Question 1.1**

Interviewer: Do you feel confident with the current state of the collaborative problem-solving assessment with an agent based system?

IP02: That's a good question. I guess, the whole construct is so messy, and I'm sure everyone who's doing research on this field would tell you that so far, there has been no completely satisfying solution to assess this construct. The problem is, I'm not pretty sure if there could be any way to assess this construct in a way that everyone would say that yes, this is the one. This is the one which is in every way satisfying. I also think that more time should have been spent on the development and on the research before the PISA decided to assess the construct, and actually assess it in 2015.

IP02: It was not at all well-prepared and well-established. I couldn't even think of too many assessment tools, either agent based or not. We don't even have too many experiences which I can tell my opinion about. Shortly, the answer is no, I don't really feel comfortable with them. No.

Interviewer: Okay. When you're saying research development should've been conducted more, do you think, is about the technical research, or in which field especially do you think that research and development should've been conducted more, because you have so many fields, you have educational, social psychological ... Now I got a worm on my tongue. The social psychological. Now I got it. Then, the technical field, and so on.

IP02: No. It's not the technical field. I guess, I'm not an expert of the technical field. As far as I can see it, everything which could have been done from the technical support, was done. It's not about that. It's much more the educational psychology field. We should have collected much more information about first of all, what is at all collaborative problem-solving, because we don't have detailed information about the construct itself. We don't have the exact structure, we don't know how the two main components, the cognitive and the social component, are related to each other.

IP02: Actually, we have an artificial construct, but we don't have actually a proper knowledge about it. This is just one side of the problem. This is a theoretical problem. Also, we did have enough experience to find out what we should treat as an optimal behavior, collaborative problem-solving behavior. How should we assign different levels to different behaviors. I could go on and go on. It's not the technical part. I'm pretty sure that if we would have ... No, if we had enough experience, again, it wouldn't be the technical part, which would give the problem.

### Question 1.2

Interviewer: Okay. It's already very interesting for me, because it's very good. Okay. Then let's just continue. That was basically the second question. Do you feel confident with the current state of the technology that is available to conduct the collaborative problem-solving students.

IP02: It depends. What are the aims, because if we talk about technology based assessment, one of the main advantages that your data set can be automatically evaluated. You don't have to do it by hand, but the computer itself does it for you. Where the problem is, the text. As you know it, the text, the human speech, still cannot be analyzed by a computer, a way that the human would do it. Still, we cannot evaluate human speech, whether it's a written or an oral speech. As long as this is not solved, this problem, the whole idea of assessing collaborative problem-solving will be, I don't know, will be somewhat damaged, because first, we should be able to do that, to analyze automatically the human discussion.

Interviewer: That is quite interesting, where I'm also reading about, general about this topic speech analyzing, because there has been some really amazing progress on it.

IP02: I know. I know about it, yes. I know that there are some embedded vocabulary. The programs can analyze based on syntactic signs or different, I don't know-

Interviewer: Actually, the analyses are not even that bad, already. The problem is, I think, rather that you have lots of languages in assessments.

IP02: That is a problem of the PISA, yes. Yeah, that's also, it's very much upon the language. For the English language, this aim I guess, it's

much closer in the future, to get realized, than for example, to the Hungarian language, which is much more complex. Much more.

### Question 1.3

Interviewer: Yeah. Good. Have you worked with other tools like this kind of tools and the CPA item builder from DIPF to create collaborative problem-solving items?

IP02: Oh yes. Actually, I haven't worked with the CBA.

Interviewer: Oh, perfect. Which other ones have you used?

IP02: I know it, because once Ingo showed it. That was I don't know, four years ago. I have never worked with it. We are doing our research based on our online diagnostic platform. It's called IDIA.

Interviewer: Okay. Is it public or is it private?

IP02: That's a good question. Yeah. I wouldn't say it's public. No, it's not. No, it's not. It was supposed ... I couldn't really tell. Is it public? Is that okay if my husband tells about this, because he knows the official version.

Speaker 3: Yeah, shortly speaking, this is an online assessment platform, you can design and create items or even questionnaires, whatever. It has a built-in item builder in it.

IP02: It's also a test system. It's two in one, actually.

Speaker 3: It's a platform for schools as well. They can register through this system, and after that, they can send us identifiers for the students from a national database, what we have here in Hungary [Country-name], and they can carry out assessments in different domains, from science, mathematics, reading, to thinking skills. We have a lot of items and other test types. You can check out our demo test. Demo test, but we have actually more, than those things. We don't have enough time to show all of them in the demo test.

Speaker 3: Yeah, this is a kind of system. Actually, it's a communication platform, it's an assessment platform.

Speaker 3: For school practice, and also a suitable system for researchers as well.

IP02: Yes.

Interviewer: Oh, okay. Oh, that's interesting, thank you very much. [crosstalk]

### **Crosstalking – Preview Question 4.2**

*Speaker 3: Shortly speaking, yes, I guess IDIA system is not compatible with the TOA system, or the TOA system is not compatible with the IDA system. In a direct way. Of course, you can build items in each system, but you have to build them from the basics. It's not just import or export or something like this, it's not like that. Not like that.*

- Interviewer:* How comfortable do you feel with the output that you have through your tool? The output and to then conduct more research on it? The outputted data that you have from your items, are they good?
- Speaker 3:* Yeah. Actually, it's a good part of this system. We have good data for research. In case of the collaborative problem-solving-
- IP02:* That's another issue.
- Speaker 3:* ... a lot of research has to be done, because this is much more complex. The items and the whole process, there's a lot of things going on during the test completion. We already did some development, it is not some-
- IP02:* You should know that ... First of all, my husband is also a researcher at the same department. We were doing this research together. He's quite involved in this CPS research. He's a much better person in this technical questions to ask, because simply, he had two years at this, which was that part [inaudible 00:19:24]. Anyway, he's mostly involved in this technical questions.
- Speaker 3:* Again, shortly speaking, it's like, yes, the data quite suitable for research purposes, if you have a normal inductive reasoning test, zero or one score, or you have a questionnaire, you can score them quite well. Actually, it can handle a lot of item types, and you can request your database in a lot of formats. We have in the download page, we can click different boxes, you can imagine. I would like to have a CSV format, I would like to have this kind of database, I would like to do only ... I only need just the scores please. I need all the data, I need the time data as fast, or something like that. You can imagine.
- Speaker 3:* There are several options how to download your data, it's not like you just get the database. You can customize your database in a way. It's quite good. In case of collaborative problem-solving, however, it was much more difficult. Actually, I would say we are still working on it. [crosstalking]

#### **Question 1.4**

- Interviewer:* The last question of this part would be, but I think you already said it, but just to get it together again, what is the most important aspect then, that the assessment of collaborative problem-solving needs, that has to be improved, that needs improvement? The most important part.
- IP02:* The most important, it would need more research background, definitely.
- Interviewer:* Okay.

#### **Question 2.1**

- Interviewer:* Okay. The next thing, where I'm digging in a little bit, because I think this is where I want to get a little bit more into details for my master's

thesis. It's the artificial characters really, so said, mostly you're mentioning it as the agents themselves, the agents in the assessment.

Interviewer: *Can you briefly describe the interaction of a student, with an agent in a current item?* Speaking from your experience, I know now that you don't use a CBA, but how is the interaction taking part in your item?

IP02: In my item?

Interviewer: In the items that you usually construct for collaborative problem-solving.

IP02: Yes. First of all, our assessment tool is yet, a human human assessment tool. We don't have an agent. Which means, it's ... I would say it's much more realistic. That's the whole problem. One of the biggest problems, with the agent based concept, that of course it's artificial. This is for a reason, that's for standardization. Without an agent, you cannot realize the standardized test environment. This is why the whole idea came up.

IP02: Yes, we use humans, and our idea is that for embedding an agent, the starting point should be first of all, to build up a human human version, and then collect many data with this version, and then based on these data, their analyses create an agent. I really think this is quite a problem of the current agent based assessment tools, that this process has not happened. Do you understand it?

Interviewer: Okay. I understand it. Yeah.

IP02: Important stages have been missed.

Interviewer: Yeah. I felt like this quite often as well. I could see PISA, you have the interact is usually only the artificial characters that have predefined, it's a little bit more modified multiple choice, actually. Predefined dialogues, and then you cannot really answer what you actually want to say, but you have to click again. I felt like this as well, that it's very often like this with artificial characters, that you should observe, of course, people in this situation. I think your approach is very interesting.

Interviewer: To have the real people, and then really conduct, how do they behave in collaboration situations like that. That's very good.

Interviewer: How can I imagine, do they all ... For example, there are three people in a collaboration situation, and three people sit on three computers, and then they just write in a checkbox with each other, just to understand how they interact-

IP02: The current version is a two people version. Yes, they are on different computers, and they are chatting. In the latest version, the communication was fully restricted. First, we started with letting them typing in messages too, but we visualized the process where we have to reach the point of completely restricting the communication. Otherwise, it cannot be automatically-

- IP02: ... evaluated, you see.
- Interviewer: How did you restrict the communication?
- IP02: You see, one usually used method is the predefined messages. Which the PISA also used, and many other assessment tools. They could send predefined messages, and also we came up with new methods, like the tests were interdependent, which meant, do you know this term, interdependent?
- Interviewer: Mm-hmm (affirmative).
- IP02: That means that they can only solve the problem together. They see different part of the problem. They actually have to collaborate to solve it. They could send actual images of their own desks, of their own screens, and ... I guess I could send you some pictures, maybe it would be easier, which we will publish. Would you like to see some pictures, maybe it would easier to imagine.
- Interviewer: Because I remember Ingo also said that you were having the research with actual people. Then, you came to, you had problems with it. That's what you were just saying, because we're first going with an actual language, and then you could not really structure language and results, right?
- IP02: Oh yeah. That was quite long ago. Which, he mentioned I guess. Then we realized that actually, that sort of problem is not at all appropriate for creating restricted communication. That is true. We chose another type of problem, the MicroDYN problems, do you know them?
- Interviewer: Yeah. I see that. Yeah. I have-
- IP02: [inaudible 00:34:20] and his team builds MicroDYN problems a lot, they have carried out many researches on these problems. Also, the PISA 2012 contained at least two third of the problems were MicroDYN problems. We also build upon this approach, this model. Only, we made it collaborative. You should also know that it was actually Samuel's team, but the girl who started this research has already left. She has also turned MicroDYN collaborative by an agent. She embedded an agent.
- IP02: In some ways, there are some similarities between her instrument and mine. She didn't start her research with a human human condition. What we are doing is that, we turned it into a human human assessment tool, and making the different variables available only for the different students. Half of them can be seen by student one, the other half by student two. They have to communicate about it.
- IP02: It's much easier that you actually are familiar with the MicroDYN. It's easier to talk about it like that. They could ask for information, send information, and the information were little images of the diagrams. You understand?
- Interviewer: Yeah.

IP02: Okay. I will send you some pictures too, which we will publish before, or after we are finished. Then you can imagine even better.

Interviewer: Yeah. Of course, I have to read about the MicroDYN. Then I also found a very interesting publication, I can't tell you which one it is, but I think it was also [inaudible 00:36:56], where he was actually writing about three different implemented scenarios of consistency, where he was playing with the consistency or inconsistency with a real world situation. That's a thing that is very ... It's actually a question that I was writing down, if there is a similar approach in the collaborative problem-solving as well. I read it within the MicroDYN publications. Now I know that you have tried the approaches.

IP02: Yes.

### Question 2.2

Interviewer: Okay. Then the next question. *How would you ideally expect an interaction with an artificial character, now we figured out problems with real world people, and with the artificial characters, and that they both bring problems.* From your imagination, what would be the best situation, the most ideal, if everything technically would be possible?

IP02: Ideal, most ideal. Yeah, that's a good question. It's hard to tell what would be the ideal ... It's easier for me to tell you my other problems. Right now, I feel that, and of course it also has a reason, that it is the agent who leads the conversation. Of course, it's the way it should be done, because otherwise, the test couldn't end. Yeah, the computer leads you over to the test, and in a way, facilitates.

IP02: What I feel that the human, the student, who takes part in this test, cannot initiate, because of the restricted communication, cannot properly express himself or herself. Of course, the ideal way would be to use free messages and to feel that the agent actually understands what you're saying. If anything was possible, then yes, I wish that he would understand what I'm saying, and would collaborate just like a human. This is something I don't know, it's quite hard to imagine.

Interviewer: Yeah. I can imagine. [crosstalking] ~~Have you heard about-~~

### Question 2.3

Interviewer: If you have to build a collaborative problem-solving item right now, what else is missing for the best possible implementation of such an item? Of the artificial character in the item. What are you missing? I already said it, but I try to narrow down, I circle a little bit around the topic. What is missing for the artificial character, if you're creating the item? The natural language or the understanding.

IP02: Yes. I don't know. It's not even the artificial character. For me, it's more like the human side. There are way to view opportunities to express yourself. You cannot express your feelings for example, you cannot use smilies in most of the cases, because it cannot be again,

evaluated automatically. While the agents are usually preprogrammed to show some kind of feelings, I don't know, for example, they give you a smiling smile, or something like that. I guess it's more frustrating from your side, than the agent's reaction.

IP02: It's not the agent part, what I have problems with, but the tool itself, in its complex way.

Interviewer: The human has actually no possibility to express anything, everything.

IP02: Oh, no, not at all. No, with the restricted communication, not at all. That's the biggest problem with the restricted communication. Still, it is necessary, because otherwise, there would be no way to automatize the evaluation. That's why I told you that this case is so messy, and it's simply, it cannot be solved in a good way, because every scenario you choose, there will be issues. There is no good solution.

#### **Question 2.4**

Interviewer: Yeah. I see. Maybe you have read the next question already, but it's actually, this is a point where I wanted to get. The next question is, how would you describe the perception of the students through the system? That's what we were actually saying, that it's not really existing.

IP02: Yes.

Interviewer: The system does not recognize any state of the student, any feeling or whatever. Then the question is, does the student behavior, such as anger, mistrust, fear or doubt, influence the artificial characters as the agent?

IP02: As long as I have heard, not yet. You must have read about it, that the agent based system so far, for CPS assessments, are these so called minimalist agents. It's only written. We did three or five predefined messages to be offered, for exchanged, which, and the data set is dynamically changing, based on the last message. It's always changing. It's highly determined, actually. As it's quite minimalist, like the name says, no, there is no space for expressing feelings, or ... No. Actually, it's much more focused on the cognitive part of the assessment, I would say, than the social part.

IP02: That's why you can also sense that it's, in a human way, in a human sense, it's not really a collaboration. It's not realistic. Although, the aim was to create a realistic part, a realistic assessment tool. There are just too many constraints.

#### **Question 2.5**

Interviewer: Okay. How does the agent or the artificial character influence the quality of the item?

IP02: Of course.

Interviewer: How?

- IP02: For example, I have ... In the PISA, there were two to four virtual agents, as far as I heard. The conversation itself can be different depending on how many agents are in the assessment tool. In most assessment tools, there is only one agent. That's the simplest way.
- Interviewer: Does it increase, the number of agent, does it increase the quality of the output, the more agents you have, the better is the result? Of course, in the end, you want to say, is the student able to collaborate.
- IP02: No. Actually, I just tried out the PISA item, which they released. My experience was that, with having more than one agent, it's even less space for you, for the human. That's just yes, an opinion. It was my feeling. Nothing research based would ... It's nothing new.
- Interviewer: Yeah. There is no correlation between the number of agents and the quality of the result.
- IP02: No, because you're the one who's thinking. The understanding. It can't improve your output, because they are only agents. Yes, in the real world, it could happen that if the group is bigger, the output could be better. You should know that it's not even this simple. It's highly upon the type of the problem. Some problems are better to solve for one person. The output won't be better if it's done by a group.
- IP02: It's usually the quite complex problems which are better to be done by a group.
- Interviewer: Yeah. Okay.
- IP02: In this case, you should really separate in your mind, this idea that the bigger group, the better output, because they are still agents.

### Question 2.6

- Interviewer: Yeah. Okay. Do you think a personalization of the artificial characters could help to enrich the item type, as in, for example, a visual personification of the agent, providing with mimics, for example. The agent says, are you sure you consider this as your question or something, and then he has, you have, are you sure a questioning phase, determined phase, something like that. Do you think mimics of the agent enrich it?
- IP02: Yeah. I'm thinking. It would be an interesting experience, and maybe it would be more motivating to go through a test in this way. You see, you said an example, again, which sounds like, I don't know, a tutor or a mentor. It's still very artificial. I don't know.
- Interviewer: Yeah. It's very static. It has smilies, they do not move at all, but they do have influence about the transported message. That's why we all use them so much. If the artificial character is supported by this, to give ... We read so much out of the faces of our partner, our communication partner, and that's one channel that is missing. What makes a personalization of a character is a visual personification, of course a linguistic personification, that is not given because the language is very

static, it's just written in facts, and it does not have, for example, rather informal language. Then of course, the characteristic personification, as in somebody who is always supporting you or somebody who's always against you.

IP02: Oh yeah. I'm not saying it doesn't sound good, but still, the main problem is that if you read through, for example, if you, have you tried out the PISA, that one problem which is released?

Interviewer: The Xandar, I think it was.

IP02: Yeah. Have you tried that out?

Interviewer: Mm-hmm (affirmative).

IP02: You can have the feeling that the messages are fully independent from each other, in a way. You don't have the feeling that you say something, and the next message will be a reaction to that. As long as you have this problem, I don't think that having facial expressions would help. Because if I tell you something, and you react to something very else, I don't care if you smile or not.

Interviewer: Yeah, that's true. I see your point. Okay.

### **Question 3.1**

Interviewer: Now I got it. To understand the structure of the communication. [crosstalking]

IP02: Right.

Interviewer: Ah. Okay. That's what I was saying. I circled around, and we already mentioned lots of things I'm asking now. Can you briefly describe the communication that is taking part in the collaborative problem-solving assessment? How is the communication taking part? What is ...

IP02: In my tool, or generally?

Interviewer: Yeah, talk about your tool. I think that makes it easier.

IP02: In the current version, it's completely restricted. Students can send predefined messages, they can send images of their own screens, about diagrams, which is another good word for that. The graph, yeah. That. Graphs, and they can also create complete sentences out of different elements. Basically, that's it. It's quite complex. Maybe in this way, I'm telling you, it doesn't sound, but it's very much. It's quite the challenge. This was one of our experiences, that it takes time to learn how to communicate. Our conclusion was, that it actually part of the problem, to find out the best ways, and to learn in way, how they should communicate.

Interviewer: Yeah. It's good you were already enriching with the multimedia content.

IP02: Yes.

Interviewer: It's already one step further. Also, from the channel back as we were saying, so far the students did not have the possibility to enrich towards the group, let's say. Except for only with dial-up. That's really great.

### Question 3.2

Interviewer: How would you ideally expect the communication to happen in the agent based system?

IP02: It would be free. Yes. Ideally, whether it's agent based or not, it should be free. The problem is again, that it cannot be currently with current technology, I know that there are some advancements-

Interviewer: But they do not really work for research.

IP02: No. This sort of assessment and these methods couldn't get valid results, couldn't produce valid results. Maybe it would mirror something, I don't say that it would be completely useful to use these kind of methods, tone analyzers or speech analyzers, based on the syntactic characters of the discussion-

Interviewer: The semantics are lost.

### Question 3.3

IP02: Yes. That's just a little part, the whole idea. Yes, the ideal way would be to discuss free. And, not only with ... The ideal way would be, from a psychometric perspective, would be to assess one person with thousands and thousands of other persons, which is again, not realistic.

Interviewer: Yeah. How? Can you describe this more? One person with a thousand people?

IP02: Sure, because to make generalizable results, this is such a case that you would behave in front of any people, differently. Actually, to produce exact results, you should be assessed with all humans in the world. Understand me?

Interviewer: Okay. Yeah.

IP02: This is one of the biggest problems of the whole concept. That's why the person who find out who is a little bit crazy, yeah.

Interviewer: Yeah. If I solve the problem once, it's hard to let me solve it again with a different kind of person. Because I already know how to solve it. I get your problem. It's interesting. Yes.

IP02: Yes. This is where-

Interviewer: I understand you.

IP02: ... the problem starts.

Interviewer: Yeah. Okay. It's already the answer to the third part, yeah. If there's anything missing for the best possible solution, for the communication. We already have that now.

#### Question 4.1

Interviewer: Okay. Now, I have answers of understanding the assessment instrument. You can also answer for yours if you can. Not on the CBA item builder, but for your item that you are using. Your instrument to create the item.

IP02: What was the beginning? Sorry?

Interviewer: Okay. I'm trying to understand a little bit how the assessment instrument is working. *The first question is, and I'm saying, since you don't work with the CBA item builder, you can answer for the item builder you have available, that you are working with, if you can answer.*

IP02: How it's working?

Interviewer: Yeah. The first answer is, if there is something missing when you think about the agent based system has ... Is there something missing, when you think about the agent based system, that a real life interaction has. If yes, what is it? It's again another approach to the same answer, like the natural language, but what else would it be?

IP02: I mentioned the feelings. Everything which describes the human human discussion.

Interviewer: You think for example, a video recording or a voice recording would help? For example, switching completely from written text, oral, typed text, towards a situation like this?

IP02: For the human who takes the test, it would be an improvement, I guess, to be able to speak and not write. Still, the reaction of the agent would not really change, I guess. It's also about the reactions which takes a lot.

Interviewer: We have feelings and reactions that are mostly the same.

IP02: Yeah.

#### Question 4.2

Interviewer: Okay. Did you have to ... That's now shifting away. Did you have to process the obtained data that your collaborative problem-solving item generates, and how practicable was it for you to handle the data in order to have research results in the end?

IP02: I think we have talked about this too. Yes, we needed to ask for some new ways to generate a data file, because first, it wasn't helpful at all. First, I asked for a PDF file, which was actually a chat history. In my assessment tool, students, while they talk, they can see their messages in different colors, and everyone has his own color. I had actually a colorful PDF sheet, with everyone's own color in the chat

- window. It was yeah, the complete chat history, but the timestamps was also added.
- IP02: I could also follow which time has everyone said what they said. That was the first one. The second one, as my husband said, was the Excel file. CSVS, which was useful because we could move the different variables whenever or whatever we want to do.
- Interviewer: Now, you are including not only the text and timestamps, but also the transported pictures that were made, for example, the transmitted pictures? Like you were saying, you're able to send screenshots and stuff like that and diagrams, so they are now also part of the data that you're obtaining through your assessment, right?
- IP02: Not the picture itself, but yes, I have some data about that too. Yes, for example, the log file shows in which way the diagram was standing. Yes.
- Interviewer: Okay, but you don't look at the diagram or stuff like that itself?
- IP02: No, it gives numbers. It gives numbers.
- Interviewer: Okay. I understand.
- IP02: I can show you a log file if you want.
- Interviewer: That would be very nice. I think it might help me to understand the data-
- IP02: You should know that it's already an arranged log file. I had to work with it a lot, to look like that.
- Interviewer: Okay. That's also part then, that the data that you get is not really satisfying to really, let's say, start-
- IP02: It was not at all. Still, I would have many wishes.
- Interviewer: Yeah. What are the wishes that you would have? I know for example, from the HCI perspective, human computer interaction perspective, there are lots of data that can tell a lot, like click, time between clicks, click interactions, [inaudible 01:06:30] and stuff like that. There is a keystroke level modeling, for example, when we have to ... We have it in the user interface part. How long it takes to fulfill a task on a website, for example. We model keystrokes needed to get to a solution, and we premodel those interactions.
- IP02: Yes. What I would imagine, yes, it would like some meta information about how long did it take to get the next reaction, which of course, you can calculate by hand, because you can do it. It would be much simpler if the system would do it for you, and you didn't have to count that much. [crosstalk]

---

**Question 5.1 – 5.4 are not part of the transcript due to anonymization and data security**

---

.....

---

**C | Final questions & Interview end**

---

No recording.

---

## Appendix A.2 – Interview Transcript 4<sup>4</sup>

### A | Introduction

---

- Interviewer: Perfect. Okay, let's start. What is very important is that I like to inform you and I would like to ask you if it is okay that I record the meeting that I just started.
- IP04: Okay.
- Interviewer: Is it okay for you? I will transcribe the interview and I will send you there transcriptions within the next 24 hours, more or less. Just in case you have something that you don't agree, where you don't want it to be in the transcriptions that I submit with my thesis, you can certainly write me an email and I will certainly get it out.
- IP04: All right.
- Interviewer: Okay. Thank you very much. I already told you about my master thesis and where it is going. We can skip this part as well and basically dive right into the interview. Am I right? [crosstalk]
- Interviewer: Okay. Then, you can see that all of the questions that I will ask you that you can maybe relate them to the PISA 2015 items that have been created. I was talking to some researchers that were not involved in PISA 2015, so I had to create a more generic approach for the interview, but in case you were involved, which you were, it would be nice if you could relate to that. Okay. Do you feel-
- IP04: Well, I need to say if you're embedding the transcription into your thesis, which is public, we can not talk about confidential item contents. Do you know which units are released, is it Xandar unit?
- Interviewer: Exactly. Xandar. It's Xandar, yes.
- IP04: Are we only talking about Xandar then?
- Interviewer: I based most of my stuff on Xandar because that's the one that I have seen as well.
- IP04: Okay.
- Interviewer: But the original audio files will not be transmitted. It will only be the written transcription of the interview. As I said, I can also strike out parts, but you can also just refer to Xandar, which makes it way easier.
- IP04: Okay.

---

<sup>4</sup> Interview Transcript 3 is missing since recording was corrupted through technology issues.

---

## B | Main Part Interview

---

### Question 1.1

Interviewer: But it's very generic. It's not really going too much into the detail of the item types itself, I guess. Okay. Do you feel confident with the current state of Collaborative Problem-solving with the agent-based system?

IP04: Well, confident is a strong term there. From my point of view we are at the very beginning of the assessment area diving into the problem-solving field. There are some predecessor domains like complex problem-solving, about which we do have a lot of information already but the Collaborative Problem-solving part is a really new domain. I think that PISA did a great job in highlighting something there, throwing a first prototype. I don't really feel confident that they tested this at a million students across the whole world, and that even policy inferences I make from the data because it's not validated actually. I do feel a bit ambivalent there. I think it's a great start, but it's not where the real ... Other instruments used in large scale assessments like PISA do have a way more high quality there than the CPS instrument has.

### Question 1.2

Interviewer: Okay. Do you feel confident with the current state of technology available to conduct Collaborative Problem-solving in general? Besides-

IP04: Sorry?

Interviewer: Besides the PISA, let's say.

IP04: Besides PISA. There I do not know about many other research groups. Some like [inaudible 00:05:05] ACT Next. I know about the things that the IIS in Memphis is working on [inaudible 00:05:15]. I mean if you are using all of the new technologies like synthesized avatars, et cetera, then I think we could have enough authenticity there in order to have valid assessment. Focusing on PISA, the technologies that I used are so basic that it's not really authentic. Yeah. I'd say that the technologies that are available right now should be quite sufficient. I mean, with respect to what you are doing with respect to what you are doing with the natural language processing part, I think the problem here is that the space of the universe of which kind of topics can pop up in a Collaborative Problem-solving space is so large that it gets difficult to have proper semantic extraction, et cetera, but I think we are on a good way there as well.

### Question 1.3

Interviewer: Exactly, yes. I agree. Okay. Other than CBA item builder, do you know or have you worked with other instruments to create items for collaborative problem-solving? If you did, what did you like or dislike?

IP04: I haven't created any collaborative problem-solving items at all. I was involved in the creation, in reviewing the PISA units for PISA 2015. There they used TAO one derivative, one branch off of TAO. As far as I saw, they used the possibilities TAO offers there. They had very complex chat maps, and I think that's all they needed for this kind of implementation. I don't think that the CBA item builder or TAO would allow for real authentic assessment like with augmented reality or something like that, or virtual reality.

Interviewer: Okay. It's quite interesting because many people are mentioning augmented and virtual reality, which I see at the very last point of a good assessment, but it's very interesting.

IP04: Sure. We're talking about the ideal thing right?

Interviewer: Yeah. I was working in an augmented and virtual reality company in Germany before. I see also how huge the mess is to create authentic stuff there. That's why for my imagination, it's still very far away.

IP04: Yeah.

#### **Question 1.4**

Interviewer: Okay. Could you name the most important aspect of the assessment of Collaborative Problem-solving that needs improvement right now?

IP04: Based on the things that, the words I know, I'd say that coming from complex problem-solving, the problem-solving part almost vanished. It's a really strong focus on collaboration, which can be nice if you are defining your construct the way it is - it is defined, but if you want to have more problem-solving things in there, you should also make this harder.

Interviewer: Okay.

IP04: Yeah.

#### **Question 2.1**

Interviewer: Thank you very much. To understand how artificial characters meant to be created for the item, can you briefly describe the interaction a student has with an agent at the current Collaborative Problem-solving assessment?

IP04: You want me to describe, for example, the Xandar item? The interaction?

Interviewer: For example, how the student is interacting.

IP04: Okay. At least in the PISA instrument, there are two ways of interaction or two types of interaction the student does with an artificial agent. The first one is communicating with them, with the fictional agents via chat or emails in some units. There they have closed range of possibilities where to meet the conversation, which is the most constrained part in my point of view. The second way of interaction, not really with the agents, yeah but sometimes also with the agents is the problem space

where they can take notes and interact with some objects, for example like Xandar with the questions that are raised by the teacher, these kind of things.

### Question 2.2

Interviewer: Thank you very much. How would you ideally expect an interaction with an artificial character in an agent based system?

IP04: Again, in an ideal world there would be also some emotional cues like some facial expressions. We would need some kind of embodied avatar there, which like I just said is really complex to accomplish in an authentic way. For example if you look at the works of Stephan Kopp there are quite a lot of good synthetic avatars produced. Also, what troubles me the most is that PISA only rewards students for the most straightforward way of interaction that you should use in order to achieve your goal. They are not rewarded for creating a stronger team cohesion. For example, if they just feedback to some person, "That was really well done." This will have future gain in the future teamwork, and this is nothing that PISA rewards in the scoring.

### Question 2.3

Interviewer: Okay. If you would have to build Collaborative Problem-solving items now, is there anything else for the best possible implementation of an artificial character, for the item itself?

IP04: Can you say it again?

Interviewer: It's sort of similar. If you would have to build an item now for the Collaborative Problem-solving, is there anything else missing for the best possible solution? You were now mentioning faces, or-

IP04: Yeah, so avatars would be nice. I mean of course there comes some more problems into play then because you have stronger, for example, gender relationship, because the avatar needs to have some gender, then it's dependent on whether the student is male or female, whether they like the face or not, whether that's sympathetic. These kind of things come into play then. This would need to be something that research and empirical evidence could suggest whether this would need to be change, or whether these changes have problems in turn. On the other hand, the step you're doing with the chat box, I think this is the very next step to take because this opens up the problem space and makes the problem space more flexible and more complex. Right now, this is really used to an unauthentic kind of way how no collaboration works at all.

### Question 2.4

Interviewer: Yeah. Okay. How would you describe the perception of the student through the system, and the other way around. How the student's behavior is perceived by the system, or how does it influence the

artificial characters. Maybe let's start with the first. What is the perception of the student through the system?

IP04: This really depends on the implementation. For example, in the work I know about from ACT Next, [inaudible 00:14:16] there they take into account a lot of information from student, like getting center information, where they are in the room. The system can adapt to all of these kinds of information. For PISA, it's really the other side of the pole, where you have so little information about the student themselves. Like only the clicks they are doing on the screen. I mean if you think of which kind of interactions are scored, this is only whether they've clicked on the link or not, whether they joined the chat first before answering any questions, or before looking into the problem space.

IP04: Then, the main interaction of course is which response did the student choose in the chat. Yeah. They are ... There is not really a lot of information taken into account. If you think of open-ended responses and text messages by the student, they could extract for example the semantics, or the problem-solving part. Also, you could look at team creation things, et cetera, social interaction. Do the students use nice words, encouraging words? Or are they kind of bashing the other team members?

### Question 2.5

Interviewer: Okay. How does the artificial character influence the quality of the item, or are there other aspects to be considered?

IP04: I think this is an empirical question that hasn't been answered yet.

Interviewer: Okay.

IP04: There is no answer.

### Question 2.6

Interviewer: Okay do you think a personalization of the artificial helps to enrich the item type? You were just saying it, but just to mention there is visual aspect for example like providing faces, or linguistic personification, talking in a rather informal language just like the approach that I do with the natural language processing, or even text to speech, or in the end even the characteristic personification?

IP04: Again, this will be an empirical question. From my personal feeling, I think this is needed if you want to have a logical, valid measurement. If you want to have some authentic setting. Auditory information would be important. There again the gender comes into play, whether they are sympathetic to the student or not, the age group. Again, makes it more problematic because students will recognize whether the voice is older than they are, or younger. This will make a huge impact on whether they perceive the other agents as being incorrect or not.

**Question 3.1**

Interviewer: I agree. Yes. Okay, to understand the structure of the communication, can you briefly describe the communication that is taking part in a typical Collaborative Problem-solving assessment right now?

IP04: You mean in the assessment?

Interviewer: Yes.

IP04: Or in the real world?

Interviewer: In the assessment.

IP04: Right now, it is really focused on the topic itself. Off-topic is not allowed at all. I think that's the main characteristic of the assessment and where it differs from real-world.

**Question 3.2**

Interviewer: Okay. How would you ideally expect the communication to happen in an agent-based system. This goes back to how is a situation maybe naturally in Collaborative Problem-solving.

IP04: Yeah, in order to be in an open-ended space, people would have a fixed amount of time for what they're doing, and vary the amount of interactions in this fixed amount of time. Right now, we don't have a fixed amount of time, but a fixed amount of interactions, which is a problem, because we are encouraging people to make X interactions instead of saying, "Well, you now have 45 minutes to go for your project. You can do whatever you want to." If you're engaged, make 75 interactions and the less engaged people will do only two interactions.

Interviewer: Yeah. I agree, it's way more natural. Some like to circle around and some just are focused on the one idea that they think is the best. It's true.

IP04: In PISA, for example, they're really forced to give answers, and we know the problems about forced choice, item format because you don't really know whether it's a valid response, whether they wanted to take this response or not. Otherwise, they wouldn't have proceeded in the unit at all, so they were really forced there. It's the kind of thing that, it's the first instrument in PISA where you don't have any missing values.

Interviewer: Yeah.

IP04: They were really forced to do anything.

**Question 3.3**

Interviewer: Okay. If you had to build a collaborate problem item building now, is there anything missing for the best possible solution to communicate within the Collaborative Problem-solving assessment?

IP04: You mean with respect to implementation? Yeah. I'd say the next step really is to open up the text messages. You will need to have some kind of semantic processing there. Maybe also some other features

would be interesting with respect to the language they're using. All this would need some engineering there. Again, the difficulty there is definitely that it's a large space in which the interactions can take place. Then, the granting more information would be necessary with respect to having some visual cues and auditory cues, which means it would need some kind of avatar, maybe even voices.

#### Question 4.1

Interviewer: Okay. Then to understand the assessment instrument itself, is there anything missing when you think about the currently-used agent-based system that a real life interaction has but the current assessment instrument does not have. [crosstalking]

IP04: Yeah. Another thing here would be for example, pauses. In PISA, you have some small delays when the student is responding, and then the delay, and then the response comes. It's a very small delay. Sometimes pauses can make a huge difference in collaboration, because then you can perceive, if you are good at social interactions, you can perceive that your partner has been thinking, or maybe they're struggling with something. You've just said, so this would be something interesting. We don't even have something like emoticons in there. This would be the very pragmatic next thing to do, because this would be very easy and these days 15 year olds would understand what they mean. Emotional cues would be something that would be missing. In an ideal world, you again, would have some facial expression instead, but this might be a next step.

#### Question 4.2

Interviewer: Yep. Okay did you have to process the obtained data of the Collaborative Problem-solving results or the data that has been produced through the CBA item folder from PISA?

IP04: Not in the primary analysis, but in the secondary analysis I did process the data.

Interviewer: Okay. How practical or how good was it for you to handle the data that was given? Did you feel that there was something missing or in order to write your report? [crosstalking] How handle-able was the data for you?

IP04: In PISA this is really straightforward, because as a national center you really get the scored items. It's just like any other instrument there. You even get the scaled information on students so that's no problem at all. I also went a step further in the secondary analysis where we are processing the process data. This means we don't only have the scored item data, but I know which response they chose, in order to see whether this can give us some information, whether some students would typically choose one distractor, but they are still highly able, something like that. I mean it's typical process data, which is very unique. You always get process ... There is no standard yet.

[crosstalking] I mean, this is not a particular thing for Collaborative Problem-solving, so there wasn't anything that was missing. I mean something you could add is in the public use file there is only the score card, so the public researchers can only use zeros and ones, and maybe twos sometimes whether the students went well in their choice, but they don't see which raw response they actually take. This is something that could be improved in the public use file. Like I said, I do have the data available.

Interviewer: Yeah. Okay. Because I'm thinking of the meta- and para-data that can be obtained through interaction by a system which of course brings way more complexity. [crosstalking]

### Question 4.3

Interviewer: Okay, where do you see agent-based systems in your daily life? If you have seen some, which I guess, or if you're actually using them, which ones are you using?

IP04: There are many on the internet when I am chatting with some enterprise as a customer, I also use Cortana in my Windows phone. I have Alexa at home. Everywhere for me.

Interviewer: Okay. Which one is your favorite one, or what do you like most about the agent-based systems?

IP04: They have some different advantages. Cortana was at the beginning quite good compared to Siri. I don't have Siri available for me. I would have, but I don't use her. I don't know why. On the iPad, I don't really ... For me, Alexa is very good. I like that, I mean I really like the hardware that's in the Echo Dot, because the microphone is really, really good. I also like the tolerance, or their understanding of the language in the sense that they understand quite a lot of brevity, if there's background noise, they still understand you. I mean, it really gives you current information, recent information if you ask for it. Yeah. That's the main thing. [crosstalking]

---

### Question 5.1 – 5.4 are not part of the transcript due to anonymization and data security

---

.....

---

### C | Final questions & Interview end

---

Interviewer: To come to an end, is there anything you would like to say or is there a topic that I didn't cover that you were missing?

IP04: Let's see. I did take some notes five minutes in advance. Yeah, something that I needed to say is that the framework in Collaborative Problem-solving is quite good. Arthur (Graesser) did a great job there. The problem is how they operationalized the instrument. If you look at

the, I don't know if you know the PISA framework grid with these three collaboration skills and four problem-solving skills. If you look at how the items are distributed among these 12 CPS skills, you will find out that they are skills that are not represented by one item. That is really awkward. Yeah, so this is something that needs to be considered. The quality of the system, it is really poor, I need to say.

Interviewer: Okay. Since I do not even know, or I have never seen all the items unfortunately. I only know of Xandar. That's what I take as the base, where I can start. I also have the feeling, I would not have any doubts, or of course I'm not really an expert to criticize the grid itself or the problem-solving framework. As far as I have seen and what I have read, it seems very good to me. That's actually nothing I put in doubt in my work, let's say.

Interviewer: Good. Thank you very much. It was very helpful. I'm very, very thankful that you could give me the time.

IP04: It was really interesting and also, I'm looking forward to seeing your final prototype then.

Interviewer: Yes. Me too. Okay.

IP04: All right.

Interviewer: Thank you very much. Have a great day and I will send you the transcription as soon as I am done with it.

IP04: Have a nice day and good luck with your work.

Interviewer: Thank you very much. Bye. For you, the same.

---

## Appendix A.2 – Interview Transcript 5

### A | Introduction

---

- IP05: Hello?
- Interviewer: Hello. Thank you very much for your time.
- IP05: It's no problem.
- Interviewer: Amazing.
- IP05: It's always good to talk to different people, and I also send my students out to interview people, so I guess it's only fair that I make an effort to try to help out other people.
- Interviewer: That's very nice. Very helpful. That's great. Amazing. [crosstalk] I've seen your work. You were also ...
- IP05: Where did you see it?
- Interviewer: PISA. Like, the publication that you did afterwards, and the extension of the framework of 2015's collaborative problem-solving.
- IP05: Yeah, I can give you a little bit of background, because it's a little bit different, because in the project where we expanded the framework with colleagues from University College London, we were pretty interested in capturing physical interactions of the students, as opposed to more having agent-based kind of assessment. So, we focused on open ended engineering tasks, where groups of two to three students needed to collaboratively solve a problem, in order for us to investigate it. It's a slightly different kind of angle of approach, but I have some experience with agent-based systems as well. Also in language-based learning, but not necessarily collaborative.
- Interviewer: Okay. Have you seen the PISA 2015 questions like Xandar and collaborative problem-solving?
- IP05: Yeah, I've seen them. I'm not that familiar with them. As I said, since my end of the work more on the technical side, while my colleagues at University College London were more on the CPS side of things. [crosstalk]
- Interviewer: Great. We have already passed the introduction. Let's say, just for you, to give you the information of how I will process or how I will analyze the interviews. I will write the transcript that I might send you within the next 24 hours, let's say. I will already suggest strikeouts of the cross talking or private conversation or whatever, that is not really fitting to the subject, and also to anonymize all the stuff that does not, or just to make sure the interviewee cannot be recognized, actually.
- IP05: Okay, no problems with me.
- Interviewer: Perfect. I want to ask you if it's fine for you that I record the interview?
- IP05: Yes, it's fine that you record interview.

Interviewer: Thank you, great. And at any time, if you feel, or you see something in the transcript that you don't want to see there, I can still get it out afterwards. Just give me this in writing.

IP05: Okay, no problems.

Interviewer: The interview will be separated in five topics. The last topic is very generic, where I have social demographic questions. They will not be part of the transcript, as well. They are just there for me to write very generic about the sample that I made. Okay, then I will have general questions about your impressions of collaborative problem-solving assessments, how artificial characters should behave in such an environment. How communication should ideally be, and what would be the best expectation of when assessment is instrument itself.

IP05: Okay, perfect.

Interviewer: Let's start right away. Most of the questions are based on PISA, but I made them more generic, so that also scientists that did not contribute in PISA can give answers. But if you ever feel like you do not have an answer, just give your general estimation about it. It doesn't matter if you were working on it or not. It's just important for me, if you have an opinion about it, that also is of great value for me.

IP05: Go for it.

---

## **B | Main Part Interview**

---

### **Question 1.1**

Interviewer: Okay. Do you feel confident with the current state of collaborative problem-solving assessment with agent-based systems?

IP05: To some extent, I feel confident. I do think that it, sometimes I think that the PISA framework and other thoughts about collaborative problem-solving framework are too narrow. Then I think sometimes, if you look at some of the work of David Williamson Schaefer from University of Wisconsin at Madison in epistemic network analysis, I think there are other ways that you can also analyze collaborative problem-solving that isn't necessarily based on how PISA did it, or complementary to it. I think sometimes when you try to standardize things, especially in education, most people have some resistance to it. If you look at the low levels of acceptance to SCORM, or other types of frameworks, there's two sides to it. There's a need for assessment, and there's also the need for the teachers to have the freedom to create the education that's required for the actual people in the class.

IP05: Of course, this changes at scale, when you have massively open online courses and other types of maybe professional training. To some extent, I feel confident, but I think it's a false confidence. I think we shouldn't just be resigned to this is the only way to do it, or this is

the best way to do it. I think it's important that we also realize that, if we quantify everything, then we take away the opportunity in education for students to experience failure without that kind of recourse. Sometimes I feel that these types of assessments create a situation that is too performance-based in some types of education, or I would say all types. But certainly, if you look at the success of design education, for instance. That requires a lot of trial and error. The same could be extended to engineering type education, or software development.

IP05: The minute we begin to create an assessment, then we're not taking into consideration what the students learn when they fail, because the only opportunity is for them to succeed.

---

### Question 1.2

Interviewer: That's great. Yeah, that's true. **Okay, so then do you feel confident with the current state of technology that is available to conduct collaborative problem-solving?**

IP05: Yeah, I feel confident with the emerging technology. I think my concern would not be with the competence, but more with the ethical and privacy sides of the current technology, in terms of, I mean in the projects I've worked on, we've had limited success at looking how students perform using artificial intelligence. But I'm not sure what would happen if this was scaled up, and became part of formal assessment. Because our models are based on small data sets, and I think a lot of times, the technology shows promise, but when it's scaled up into the real world, it begins to have some challenges. In terms of agent-based interaction, there's been a mixed response to customer service based agents. I imagine there is a mixed response for students to interact with agents as well.

IP05: I think the technology is there, and certainly outside of education shows a lot of interest and promise about really talking about the ethics of privacy, of what types of data the system collects or the models are based on. I think the technology is progressing, and certainly if used in the right way, would be beneficial to collaborative problem-solving assessments.

Interviewer: Yeah, I totally agree. That's a thing that I also had to face. For now, I switched to a very, let's say good platform concerning data security, and personalized because I'm working with dialogflow, but I did not find a perfect fit in the open source product.

IP05: It's very challenging to create that kind of open systems, for sure.

---

### Question 1.3

Interviewer: Yeah, definitely. [crosstalk] **Okay, other than, have you ever seen the CBA item builder? The tool.**

IP05: No. Only when I looked it up, after you sent me the questions.

Interviewer: Okay.

IP05: I haven't really worked with this tool. I've worked certainly with other tools, and learning management systems, and developed systems of, similar to this for mobile learning. I think as I said, the problem is always that you create a model of how you want to solve the problem, and this model is sometimes hard for other people to adopt. And so that CBA item builder is certainly nice, but there are similar type tools that allow you to make SCORM compliant modules, or even more simple tools like Google Forms, or learning management systems that do similar type things. I think that I guess, as I said, it's always this kind of other person structure. Oh, this is some structure that's hard to sometimes get used to, especially in the context of teaching in Sweden, where we have a lot of freedom as educators to design the courses the way we want to. So that in general, that's not how these tools are designed.

Interviewer: Yeah. Which platform do you use, then? It's a little bit off-topic. [crosstalk]

---

#### Question 1.4

Interviewer: Yeah, exactly. **Okay, if you could name the most important effect of the assessment of collaborative problem-solving that would need an improvement now, what would it be?**

IP05: I go back to when I started with this, just making space for the creative failure, and reflection. I think we talk a lot about reflection and action, or flow and other terms of how experts work, but I think a lot of the times, we don't create ways to assess what the student's real experience was unless we ask them to write a reflective essay. Even then, in any type of collaborative problem-solving situation, it's also how the group interacted, and perceptions of the different people that we don't always look at in terms of assessment. Because I think that school is one of the only places, at least at university, where you can experience some type of failure and learn from it.

IP05: Because it's very hard to do that in the work situation, or in your personal life, but how do you train someone to deal with challenges, and failures, and miscommunication? That's sort of what we try to do when we do collaborative problem-solving at the university, or other education at other levels, but it's not that easy to design for it. Sometimes in math education, we do what we call productive failure. We, in order to have the students reflect on a bad strategy in order to learn the right strategy. We do that a lot when we teach physical computing, where we ask the students to make something that actually doesn't work, so that they understand that they can push the limits. But I think it's very hard to build a system that assesses that.

Interviewer: Yeah.

- IP05: In many different domains, in different subjects. I think that's kind of the needs improvement, is what do we actually want from the collaborative problem-solving situation? That's besides the standard 21st-century skill craft. We want people to learn to be better communicators, and better team workers. That requires failure rather than success, because you don't learn anything if you're in a group, and you do everything, and it's great. [crosstalk]
- Interviewer: Right. For example, I was working in many startups, and that's where I experienced the opposite. We were doing a mistake, we were doing failures. We got, of course, lots of feedback from investors, etc. when it was not going right or perfectly, and then we were always sitting together, and we were reflecting. Okay, what was happening? Why was this stage not going perfectly, and how can we do it better the next time?
- IP05: That's, I think, something that education has become really bad at, especially at the university level.
- Interviewer: I agree.
- IP05: That's something that I think we need to build into these, the design of collaborative problem-solving activities is that startup experience of finding that plan A is never going to work, and that it's plan M or plan L that actually keeps the company alive. It's very hard to teach that, I think because, I don't really know why. We've been struggling with it, and it gets harder and harder to teach it as well, because students aren't willing to invest so much time sometimes in doing things, they just want to do the least amount of work to pass.
- Interviewer: Yeah, exactly.
- IP05: I think that's the challenge of collaborative problem-solving kind of activities, is that you're forced into a situation that pushes them beyond what they really want to do, and designing that and assessing that is really difficult.
- 

### Question 2.1

- Interviewer: Okay. **Then to understand maybe how artificial characters should behave. Could you maybe briefly describe how the interaction of a student with an agent is at the current assessment? And then the next step, of course, how it ideally would be. But how you see it, at the current assessment.**
- IP05: I think a lot of it is certainly chat-based kind of assessments and interactions. Occasionally, of course you have artificial characters. In the case of some of this virtual internship stuff from the University of Wisconsin Madison, they can actually be actors as well. So, in the situation, you could have a video from someone playing the role of a scientist, or a politician, or an environmentalist. In a sense, it's an artificial character, but it's not necessarily an NPC type character. I

think the majority of it is text-based. You look at it emotionally, with agent-based learning and mathematics, so in that sense it's teaching somebody something else. You know, how do you solve this problem? Then you respond, and those are the steps that you do. You get feedback.

---

### Question 2.2

Interviewer: Okay. **How would you ideally expect the interaction to be with an artificial character and an agent assistant? How would this work?**

IP05: Yeah, this is a little bit tricky, and it gets a little bit to your question six here, but I think in some ways, sort of like a more game-based kind of experience, where you interact with characters that have some depth. That could change, and that could respond to you. It doesn't necessarily have to be a high resolution kind of 3D character, it could also be a text-based character as well, depending on the context. But some kind of richness to it, so that in a sense, is a little bit of suspension of reality, sort of like you're playing a game or you're seeing a good movie.

IP05: You know that it's fake, but at the same time, you're interacting with it in a way that is pushing you. I haven't really thought in the sense of like, what kind of form it should take, but more the action that actually responds to you, and responds to other people differently. So in a sense, you begin to have some type of perceived relationship with it, based on the input that you give.

---

### Question 2.3

Interviewer: Right, yes. That sounds great. You would have to build a collaborative problem-solving item now. **Is there anything else missing for the best possible implementation of the artificial characters for such an item?**

IP05: No. If there's anything missing, I guess I kind of fall back on the privacy or the ethics side. How do I balance, what's relevant? What do I want to share with the system that's assessing me, and what do I don't want to share? At least as an adult learner, I think many people have different patterns of interacting with these artificial characters. Then to what extent do you analyze that in terms of analytics? Do you find out that 20% of your students are obsessive compulsive disorders, or are using your chat bot at three in the morning instead of during ... I think there's some kind of way that you have to, if you create an artificial character that you have a relationship with, there should be some kind of privacy to some extent.

IP05: Like, how much do I want to share this, so that me, as a learner is in control? Of course, it defeats the purpose of the assessment, but on the other hand it's a matter of trust.

Interviewer: I agree, yeah.

IP05: This sort of gets, I guess, into the next question.

---

**Question 2.4**

- Interviewer: [crosstalk] **How would you describe the perception of the student through the system, as it goes into the next question? How does the system actually interact, or see the student? How does the student's behavior influence the artificial character?** That's right, yeah.
- IP05: I think there's a need to be transparent with students, so that they actually see how they're interacting. So they can reflect, is there a better way to interact? There's a lot of research that shows that most learners don't use these tools anyhow. They don't look at why they're not performing well, because they know they're not performing well because they are working, or spending time on their social life, or other issues. But I do think that people today are pretty familiar with computer games, and they deal a lot with artificial characters in those games.
- IP05: Most games are much better than agent-based systems for education, and so there is a perception of, this is crappy educational junk I need to do to go through school, just like all the educational software that we used throughout our own school career, because the budget is quite different between making a game and making educational software. I think there is, of course, anger and mistrust, or I would say maybe even frustration when students use these systems. Especially the ones that are simple, where it's like, "Good job," or, "Bad job," or, "Can you explain that again to me?"
- IP05: I think it's really important that we look at this from the usability standpoint, because I think a lot of times, people develop education software without thinking, what do the students actually think about it? We develop because we think we're an expert, and this is how it should work. It's almost wrong. I mean, it's definitely wrong. I think it's really important that we have to recognize that we're never going to make our artificial characters as engaging as the newest game character. And so, we have to find a way to balance that out in terms of what benefits or added value do students get out of using this system? Because I mean, why would you waste your time using these systems that are frustrating for everybody?
- Interviewer: Yeah, there is definitely a ... engagement is not really great in most systems, especially if we all, all students are with computers, they're playing games a lot, also in their private time, that's also a feeling that I got between what we have in educational systems and educational software towards, let's say the real world, or whatever I do the rest of the day, or when I'm gaming, or whatever. There is a huge bridge, or unfortunately right now, there's not really a bridge, but there is a huge gap.
- IP05: It changes a bit in performance-based simulations for first responders or medical education, and that stuff where it becomes much more

game-based. Then it seems to work better if you're trying to create a situation where first responders are responding to a situation. You can do that with simple game mechanics, and even simple graphics. It does provide some type of stuff, but I guess the big question would be, do we need to have an agent first?

Interviewer: If it's not an agent, what else would it be?

IP05: Yeah, but it could be a series of submissions that are assessed and sent back automatically. You're not necessarily interacting with the character, but you are submitting something that's being assessed automatically, and sent back. You know, "Your essay doesn't really contain all the elements that are needed, please resubmit."

Interviewer: Okay.

IP05: More like natural language processing kind of aspect, where there are systems that do that. Like in science learning, you submit a scientific report and it gives you feedback. Then that involves a lot of work on modeling systems. The system gives you feedback if you don't have the right kind of, let's say predators and the right type of prey. You get feedback because, of course, only your predators die, for instance. Or they eat all the bunny rabbits.

---

### Question 2.5

Interviewer: That's cool. **Okay, how do you think the artificial character has influenced the quality of the item? Or are there other aspects to consider?**

IP05: I think there's a lot of, people respond to artificial characters like they respond to characters in games, or books, or novels, or comic books. I think there's some research by, now I forgot his name, a researcher at MIT who looks at gender and racial issues in avatars. I think people respond a lot more to it than, there's not enough research on it. I think it's a very tricky thing to design these artificial characters. I mean, maybe not for younger kids in the same way. But I mean, there are other influences like gender or race that could influence children's perception of themselves by poorly designed characters. I think that it's quite tricky to design these characters, so I guess that leads also to number six.

---

### Question 2.6

Interviewer: **Exactly Do you think the personalization of the artificial character would help with the archetype? Either, like you were saying, already 3D characters, or like talking in slang, or for example have a characteristic personification? Like being very nice, or very bad, let's say.**

IP05: I think sometimes it could help, for sure. I think having sophisticated artificial characters will enrich the experience, in terms of initial personification. Certainly linguistics and characters, you know? This is

all kind of culturally bound. What might be acceptable in Europe may not be acceptable in the United Kingdom, or the US, or South America, or India. It's a little bit tricky there, to some extent. I think the ability, maybe for having people to change it, as well. Maybe I want to have a low kind of resolution version, because I don't want to be distracted. Maybe I want to shift the linguistics a bit, that could be an interesting way to deal with it. Maybe I don't want to be distracted, or I don't have a powerful device so that it wouldn't work.

IP05: It's little bit of a trade-off, right? Because it also depends that people have the right level of technology, which isn't always the case in education.

---

### Question 3.1

Interviewer: Yes. Okay, great. **Then we come to the next part, which is to understand a little bit more about structuring communications in the system. Could you briefly describe how you see the communication that is taking part in a typical collaborative problem-solving assessment?**

IP05: The work that I've done, as I look at it, I look at what we would call the engagement of the different students as individuals, and also within the group. Simply put, are they all engaged in communicating about something in-hand, or are two working on it, and one distracted, or one doing another task? This task could be part of the larger task, or they could just be disengaged. Or are they all disengaged? I think in that sense, we look both at the communication in terms of the verbal communication, in terms of who's talking and what they're talking about, and also the physical interaction part to try to understand what's happening in terms of that.

IP05: Because in some other research, you can see that sometimes students may not seem to be physically engaged in the collaborative problem-solving activity, but they actually are quite engaged in terms of sharing information.

---

### Question 3.2

IP05: I think it's both the physical side, and of course the more oral and verbal communication sides that are important to look at. Most of the time, people don't look at the physical interaction. They just look at the text, or oral communication.

Interviewer: Excuse me?

IP05: Go ahead.

Interviewer: **This is rather the ideal expectation that you would have.**

IP05: Yeah.

Interviewer: That as well linguistic, as also physical behavior could have an influence or impact.

- IP05: Yeah. If you look at the physical side of things, teaching like, let's say middle school students, physical computing. You teach one group how to make a noise through a speaker, right? Then students will come over and see how they make the noise and go to their group. Then there's this kind of physical activity that's not really tracked, necessarily in a lot of research. I think this kind of collaborative problem assessment, although it's more about activity design rather than assessment, is something to take in consideration as what's happening physically in the group, or in the larger space of the different groups. That would be my ideal kind of communications size, is what's also happening physically. Is that relevant?
- Interviewer: I guess that's something that you were approaching in the publication that you were writing.
- IP05: Yeah.
- Interviewer: Yeah.
- IP05: Because I think sometimes it's unrealistic to analyze the dialogue, especially if you are doing something with your hands as a group. You're not interacting via text, you know? You're communicating. How do you assess that without them having to record it, and analyze it, and that still is beyond the computational powers, at least in the average classroom.
- 

### Question 3.3

- IP05: But the physical interaction, course, is there. That's a little easier to do and less computationally intensive. **I think it sort of answers question three as well, where I think that we need to think about communication from a larger perspective, beyond verbal or beyond textual, but also like the physical interaction.**
- IP05: How does that relay communication? Maybe we have someone who doesn't interact so well with people or doesn't communicate, but does a lot of work. We might have the opposite situation where you have someone that talks a lot and never does any work, but that person generally does much better in these assessments than the non-verbal person. I think we have to think about, what are the models we create, and are we just building in our own biases by saying that people that communicate more or better, or performing better than people that aren't? Then it becomes a real tricky bag of things.
- Interviewer: Yeah, I agree. That's true. [crosstalk]
- IP05: On the other hand, there is a lot of research done by people, like organizational research where they look at how innovation happens, or how people move around office spaces using these smart badges. I think there's opportunity there in education that we haven't really looked at yet, in terms of collaborative problem-solving. Maybe we don't need to track the whole body, but we can track who's talking to

who, when, who's standing with who, who outside of the lab is working together or discussing things. I mean, I think it's easy to collect a lot of data, but garbage in is garbage out. I think that a mistake of the engineer, you know, we think we can collect everything and make sense of it, but the reality is quite different.

---

#### Question 4.1

Interviewer: **Okay. Then to get to the next part, to understand the assessment instrument itself, do you think there's something missing when you think about the currently used agent-based system, that a real-life interaction has, but the current assessment does not have? You see it's repetitive, but ...**

IP05: Yeah, I think this is really about that. For me, it's the physical interaction, the interaction, the macro space. It's what's happening in the whole classroom, it's also interesting in terms of this situation. The way that I imagine collaborative problem assessments is in real classrooms with groups of students, doing activities together on 10 different tables. But what is happening between in that space, and some of the work that we did, we tracked what the mentors were doing. When you see that you don't always help the students that need the help. How you overcome that?

IP05: There's some work that Pierre Dillenbourg did with this, it's like in EPFL where they have the big physics class, and made this, I guess it's like a lava lamp. If you need help as a group you hit it, and then first it's green, so the TA knows that you need help. Then after a certain amount of time it turns yellow, which says I haven't been helped yet, and then it turns red after a certain amount of time. I think something simple like that, in a real-life classroom makes a difference because it actually gets students the support that they need so they can get over their problem.

IP05: I think that, if you have agent-based systems, you have groups of students working with these different agents. The agents should be talking to each other to try to say, this is where we think the class is, but this group is performing well, this group needs more support, this group isn't really doing anything. I would think that what's missing is, if we go to an agent-based system, how do these agents talk to each other to get a bigger picture of what's happening in the environment?

---

#### Question 4.2

Interviewer: **Okay, that's great. Then to the next question. Did you have to process, obtain data of collaborative problem-solving? Or did a collaborative problem-solving item generate, and how practicable was it for you to process the data that was conducted with quantitative research?**

IP05: Well, I think in the beginning it was hard for sure. A big project that we completed more than a year ago, we're still processing the data. The

hardest thing was trying, since it was a very exploratory type of project as opposed using a tool like CPA or other tools, it was trying to make sense of the data, and then trying to find possible ways to first process it, and then to analyze it. Then once we did that, which took a long time, then that's become a little bit easier because now we have sort of a framework that we developed, a set of different tools to develop it.

IP05: It becomes easier that with different types of technologies such as unsupervised deep learning, where you can basically, once it knows who's who, it can create some type of score of the collaboration. We haven't tested it outside of the data yet, so I think it's a huge thing, but I think the problem is in this project, we never set up what was the real question. So the questions developed over time, which makes it much harder to handle the data. On the other hand, it gives you the freedom to find new patterns. I think there's always a trade-off with this type of quantitative research, just like there is in qualitative research, where you have to analyze all the transcripts, and what everyone says. I think that in any type of qualitative or quantitative ethnographic research, it's always a huge problem.

---

### Question 4.3

Interviewer: Yeah. That's true. Okay then, we already come to the last question of the main part. It is, where do you see agent-based systems in your daily life? And if you have seen some, or if you are actually using them, which ones are you using, and which ones do you like the best? What pleases you the most?

IP05: Well, sometimes I like to learn that you can dictate things to Siri, through using an iPhone, or Apple's stuff. But in terms of getting information, I think like Google works quite well in terms of voice-based searching, navigation and stuff. I think those are the primary agent-based systems I use in my daily life. I wouldn't say, it's not a large part of my life. I don't have Alexa, sometimes it's easier to use your thumb than it is to depend on the system giving you what you want. But I do see, in the future that it becomes interesting, and we go back a little bit to these more agents where you could program them to do something.

IP05: For instance, run my dishwasher when electricity is cheaper, or something like Google's Nest, that can control a thermostat in your house. Or an agent system that could support your driving, some of the driving assistant tools that are out there if you can afford to have an expensive car. I think what's nice about it is that it removes the interaction from visual, and touch, to voice or to some kind of pattern, your daily pattern. But I think it's also sometimes a little bit creepy.

Interviewer: Yeah. Have you tried to personally ... Sorry? Why is it creepy?

IP05: Go ahead. Well, who knows? Nobody likes to think that they fit into a certain category, like the way that Netflix recommends things to you. I mean, we all think that we're individuals, and we shouldn't fit into this

category. The realization that we fit into a category, sometimes I think it unsettling.

Interviewer: Yeah, that's true. Have you tried to personalize your environments? For example, what I do, I have lots set up in my chat environment as Slack or Telegram, and I integrated many of the services, let's say that just help me, give me updates on traffic and all these kinds of things. Have you tried this?

IP05: Yeah, sometimes, but not as much as I would like to. Sometimes, when you feel like the system breaks down, it's more frustrating. Like with the truck traffic, or public transport information. But yeah, I have thought about it. I haven't really done it so much. Probably just too lazy.

Interviewer: It takes a little bit of effort to set it up initially, and it doesn't really go with, I think it's a little bit more that you have to actually initiate it. Apple and Google do not do, you actually tell it to tell you something, and then it will give you an answer. I think it's the other side of approach that I, for example, like a lot. I initiate and I want to know. For example, I have the automation and updates on all my git repositories, so I know what's going on, [crosstalk]

---

### Question 5.1 – 5.4 are not part of the transcript due to anonymization and data security

---

.....

---

## C | Final questions & Interview end

---

Interviewer: Yeah, okay. **Then we have reached the end of the interview, which was very, very nice, very inspiring. I already say thank you, but did I miss a topic that I should have covered?**

IP05: There is one on topic four, that you hadn't grayed out, on the poor validity of results. I just wanted to say, I think at least in the learning analytics field, we're aware of that. I think the hardest thing we face is, how do we, in a sense, even though we share our data sets to some extent, our work is never really re-created and tested in terms of validity. This is true in many different domains as well. I think that there is a big problem in terms of the fact that other people aren't testing the systems that I built, even though down the road they're not able to test it, or analyzing the data to make sure that it is actually working, and we didn't purposefully or non-purposefully corrupt our own data to get results for publication.

IP05: I think this is something that's really hard, because decisions are made without assessments that affect people's lives in terms of saying, this person is going to end up in a vocational track, or this person is definitely headed to the academic track. That may not be true, but it

may be based on a faulty assessment framework or a bad algorithm that you licensed, that you don't even own. I think that's really, and the privacy and ethics of it. I think there's something to be said about that. I mean, who controls the data that the students generate? Is it me? Is it the university? Is it the school? Is it the company?

IP05: Where is it stored in relation to the PDPR? What happens if my data sets can predict that I'm going to end up not being a functional part of society? Does society make an intervention? It's all that type of stuff, but sort of around this idea of agent interaction. How much will the agent know about you? Who controls that? I think that's all I need to say on that.

Interviewer: Right, okay. **Is there a closing word that you would like to say?**

IP05: I think I like the idea of collaborative problem-solving and agents. I think it's something that is also interesting to think a lot about, especially if you look at massively open online courses, where you take this course and you never know who's really out there with you. As opposed to when you have this face-to-face type of communication, that you can read my face, and I can read your face, to some extent. But when you do a MOOC, it's very hard to get that. What you see is, there is a lot of hype about MOOCs changing education, and it never really happened because there isn't that type of connection. Possibly agents could be one way to improve people's experiences, and give some type of feedback and interaction. I think that's it.

Interviewer: Okay, great. Thank you very much for your time. [crosstalk]

---

### **Appendix A.3 – Interview Analysis**

---

*The interview analysis is an exported excel file, and extends the documents format, therefore doesn't fit into this document. It can be obtained online (<http://770695-2.web1.fh-htwchur.ch/interview-result-matrix.pdf>) or from the author on request via mail ([hanna.kummel@gmail.com](mailto:hanna.kummel@gmail.com)).*

## 10 Appendix B – Technical Documentation

### B.1 Software development plan

#### B.1.1 Stakeholder Requirements

*The purpose of the Stakeholder Requirements Definition Process is to define the requirements for a system that can provide the services needed by users and other stakeholders in a defined environment (IEEE, ISO; 2008).*

- The stakeholder requirements are described criteria collected throughout this thesis and thereby are conducted in a valid and comprehensive manner.
  - Implementation of natural language
  - Implementation of interactivity
  - Implementation of strong characteristics

#### B.1.2 System Requirements Analysis

*The purpose of System Requirements Analysis is to transform the defined stakeholder requirements into a set of desired system technical requirements that will guide the design of the system (IEEE, ISO; 2008).* System requirements for this platform are defined as follows:

- The system has to be a conversational interface<sup>5</sup> that is accessible to a certain group of people.
- Access to the system can be given through providing account details by the author. Furthermore, the components have to be published and accessible to the public.
- The conversational interface simulates a pre-defined scenario that resembles a collaborative problem-solving scenario named Xandar, published by PISA 2015.
- The implementation of a triologue<sup>6</sup> has to be possible in order to expose the test-taker to an equal complexity as in the original model, namely the Xandar unit (see previous bullet point).
- The implemented content may be integrated in a reduced extent but must still demonstrate the fulfillment of criteria of the stakeholder requirements.

#### B.1.3 System Architectural Design

*The purpose of the System Architectural Design Process is to identify which system requirements should be allocated to which elements of the system.*

- a system architecture design is defined that identifies the elements of the system and meets the defined requirements;

---

<sup>5</sup> E.g. a chat client such as Slack, Telegram, Facebook, etc.

<sup>6</sup> Conversation with three people.

- the system's functional and non-functional requirements are addressed; (IEEE, ISO; 2008).
- The system has to consist of two main components: A conversation-interpreting software product<sup>7</sup> that serves as language processing unit<sup>8</sup> and a graphical user interface that serves as frontend to the end-user. Furthermore, a backend to host the software components is necessary.
- Within the conversation-interpreting system<sup>9</sup>, the implementation of intent-based conversational logic<sup>10</sup> must be implementable. The conversation interpreting system may not interfere or be in conflict with multiple agents.
- Within the graphical user interface representing the frontend of the system (namely chat client), the implementation of conversation-interpreting software components<sup>11</sup> (namely agents) must be possible. The creation of group-channels<sup>12</sup> or similar constructs must be possible.
- Within both components (chat client and conversation-interpreting system) there must be the possibility to exchange text messages, trigger events, implementing interactive components.

#### **B.1.4 Implementation**

*The purpose of the Implementation Process is to realize a specified system element (IEEE, ISO; 2008).*

##### **B.1.4.1 Software Architectural Design**

*The purpose of the Software Architectural Design Process is to provide a design for the software that implements and can be verified against the requirements.*

- A software architectural design has to be developed and baselined, that describes the software items that will implement the software requirements.
  - Components: Implementation of two artificial agents and one interactive component that navigates the test.
  - Trialogue between one natural person and two artificial agents. Supportive functionality through the interactive component.

---

<sup>7</sup> E.g. platforms like Chatfuel, Kitt.io, Dialogflow, etc. that allow users to "create" a chatbot

<sup>8</sup> A software component that can interpret natural language

<sup>9</sup> Conversation interpreting system: The service provider that allows the creation of artificial characters

<sup>10</sup> A concept of language structure where e.g. an agent's action can be triggered by a user's action

<sup>11</sup> Conversation interpreting software component: e.g. an agent that can interpret natural language

<sup>12</sup> Group-channel: a chat window where more than two participants can talk.

#### **B.1.4.2 Software Detailed Design**

*The purpose of the Software Detailed Design Process is to provide a design for the software that implements and can be verified against the requirements and the software architecture and is sufficiently detailed to permit coding and testing (IEEE, ISO; 2008).*

- Construction of the conversation flows<sup>13</sup> has to be carried out before construction
  - A conversation flow represents one item of the Xandar Unit.
  - Five or more conversation flows have to be implemented for demonstrative purposes
- Construction of the interactive component is part of the conversation flows.

#### **B.1.4.3 Software Construction**

*The purpose of the Software Construction Process is to produce executable software units that properly reflect the software design (IEEE, ISO; 2008).*

- Implementation of the conversation flows into conversation-interpreting system.
- Implementation of interactive components concept into conversation-interpreting system.

#### **B.1.5 System Integration**

*The purpose of the System Integration Process is to integrate the system elements (including software items, hardware items, manual operations, and other systems, as necessary) to produce a complete system that will satisfy the system design and the customers' expectations expressed in the system requirements.*

- Rolling out software components to an accessible front-end system
- Prepare system and test-accounts
- Enable the interpretation of necessary interactions via OAuth and requests in chat client.

---

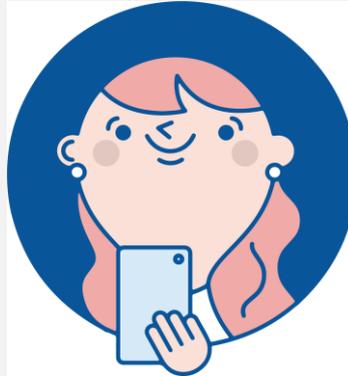
<sup>13</sup> Conversation flows: concepts of a possible conversational procedure

## B.2 Personas

### Persona 1:

Pupil with strong empathy for others

#### Picture:



#### Fictional name:

Anna

#### Demographics:

Anna is 15 years old

Has no boyfriend but many friends

Lives with her mom and dad in a small town

Is in 9<sup>th</sup> grade at a high school in the U.S.

#### Attributes:

Attributes that form Annas character:

- empathic
- sensitive
- intuitive
- ambitious
- extroverted

#### Goals and tasks:

Spends her free time with reading, meeting friends and doing her homework conscientious

#### Environment:

Is related to “Generation Z” and thereby loves her phone and social media. She lives in a house in a calm neighborhood and both parents have a job with an average income of \$60.000 each per year.

Table 4: Persona 1 - Anna

**Persona 2:**

Pupil with boorish behavior

**Picture:****Fictional name:**

Kevin

**Demographics:**

Kevin is 15.5 years old

He has no girlfriend and some friends

**Attributes:**

Attributes that form Kevin's character's character:

- disengaged
- careless
- unmotivated
- stubbornly obstructive
- unwilling to cooperate

**Goals and tasks:**

Is spending his free time with gaming or watching tv. He prefers to meet his friends online when gaming. He's missing in school quite often because he forgets time when being in the middle of a game.

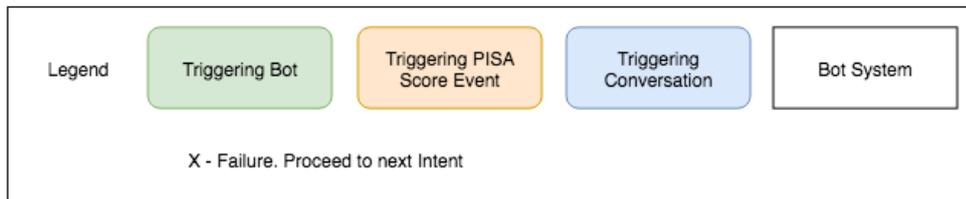
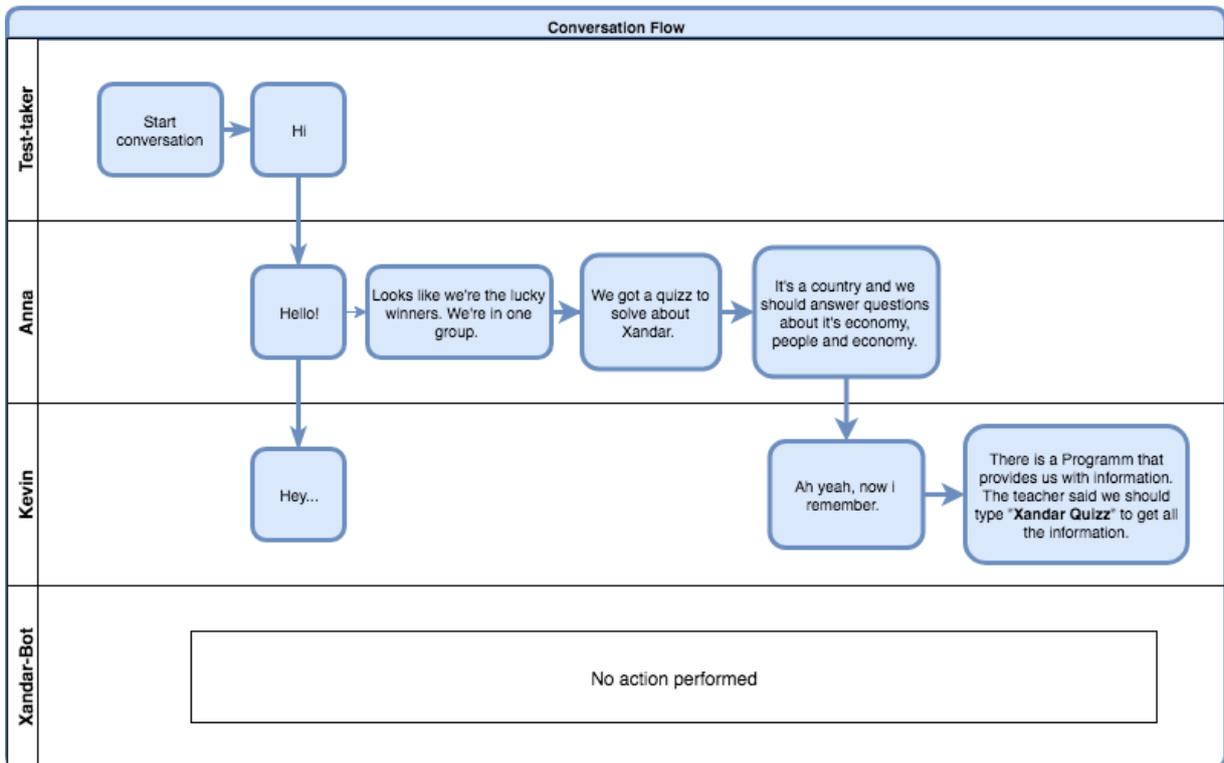
**Environment:**

Is related to "Generation Z" and thereby has a strong relationship to technology. Is not interested in social media but into technology that serves his hobby: gaming. He lives in a small apartment with his mom in a apartment block. His mom has an average income of \$30.000 per year.

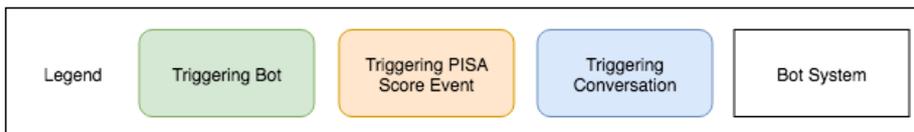
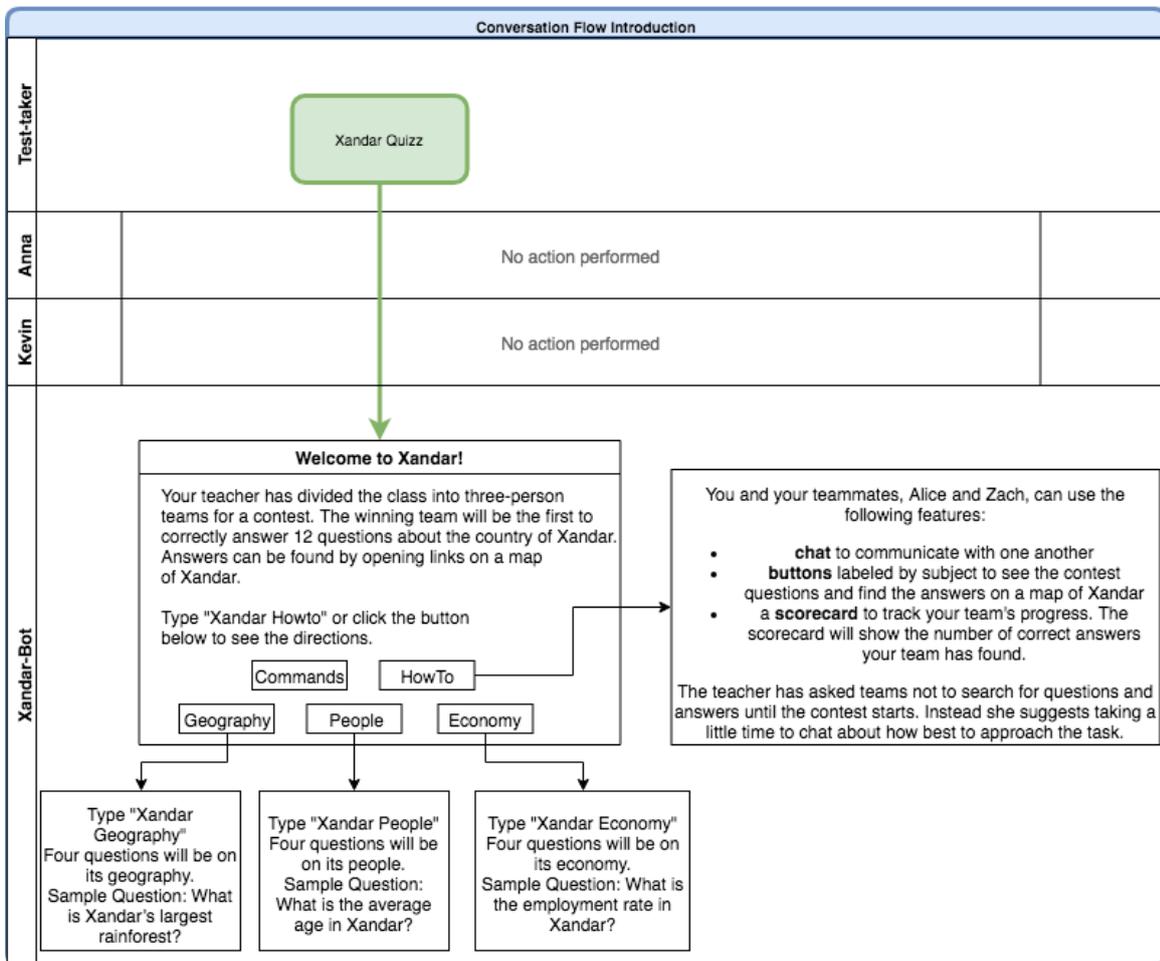
Table 5: Persona 2 - Kevin

### B.3 Conversation Flows

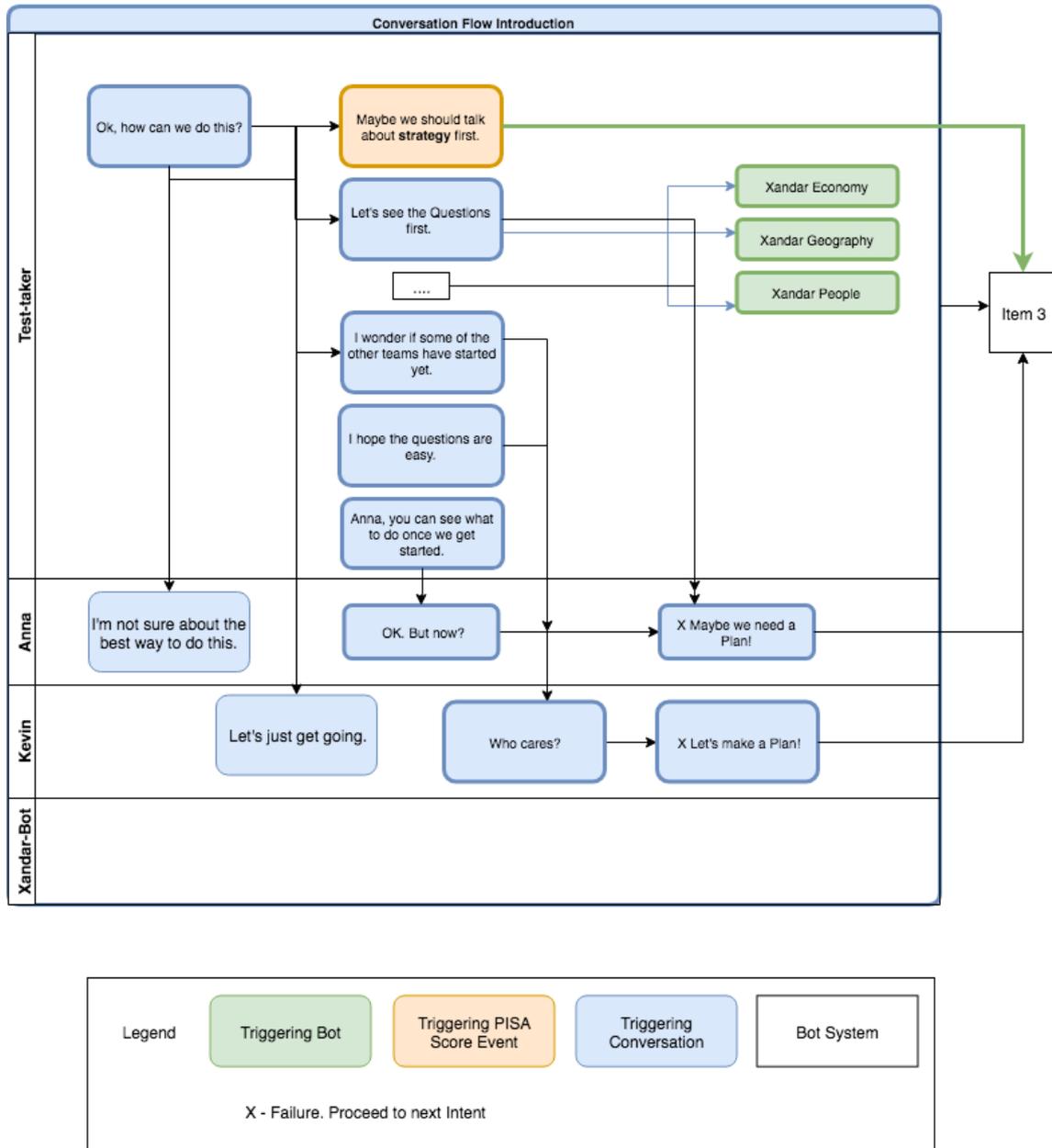
#### B.3.1 Item 0 Part 0: Start Conversation



### B.3.2 Part 1, Item 1: Following Directions

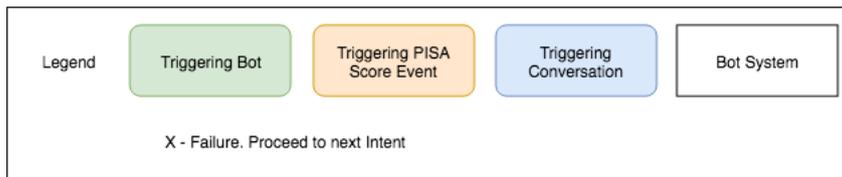
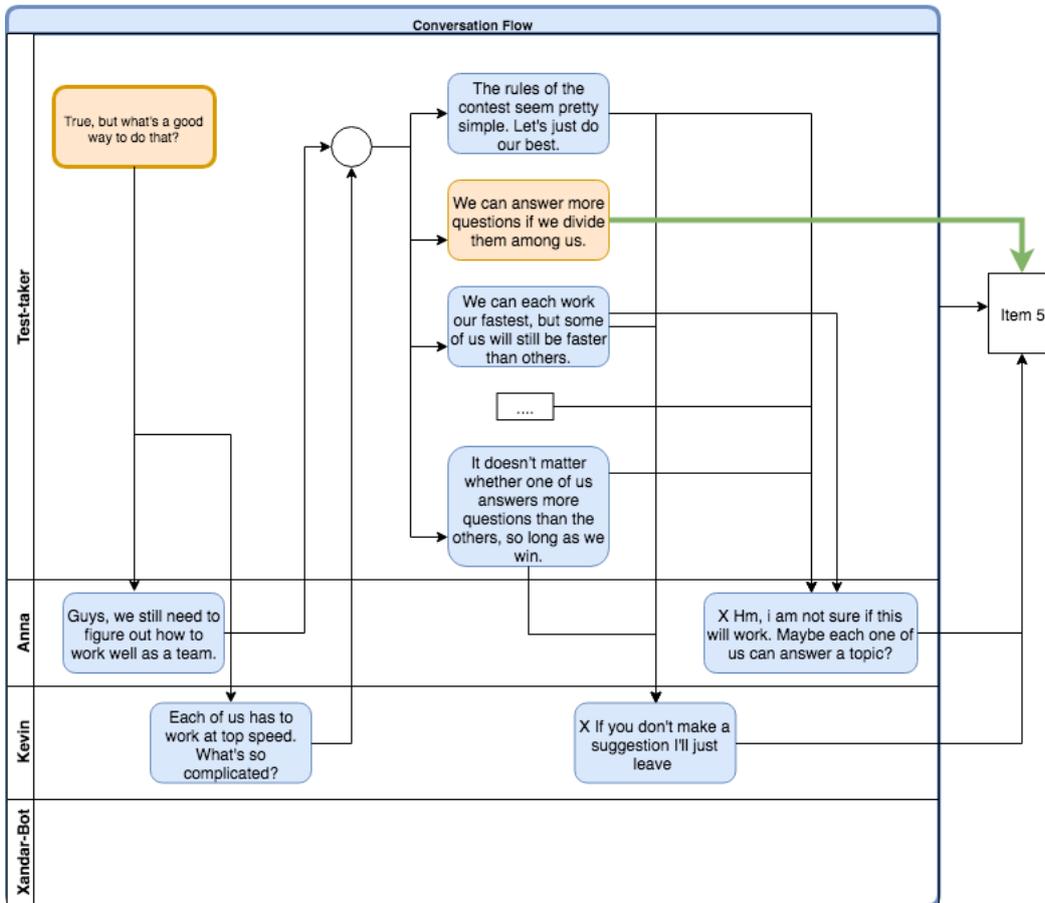


### B.3.3 Part 1, Item 2: Understanding the Game





### B.3.5 Part 1, Item 4: Agreeing on a Strategy



## **B.4 Set-Up Test Environment**

### **B.4.1 Own slack Workspace**

1. Login to your own Slack Environment
2. Add Bots to your Slack Environment
  1. Add Anna
  2. Add Kevin
  3. Add Xandar Quizz Instructor
3. Create Group with Anna, Kevin and Xandar to start
4. Have fun solving a Quizz with them!

### **B.4.2 Demo Slack Workspace**

1. Create account for abasco.slack.com
2. Anna, Kevin and the Xandar Quizz Instructor are already installed to this workspace.
3. Create Group with Anna, Kevin and Xandar to start!
4. Have fun solving a Quizz with them!



## Bisher erschienene Schriften

Ergebnisse von Forschungsprojekten erscheinen jeweils in Form von Arbeitsberichten in Reihen.  
Sonstige Publikationen erscheinen in Form von alleinstehenden Schriften.

Derzeit gibt es in den Churer Schriften zur Informationswissenschaft folgende Reihen:  
Reihe Berufsmarktforschung

Churer Schriften zur Informationswissenschaft – Schrift 1

Herausgegeben von Josef Herget und Sonja Hierl

Reihe Berufsmarktforschung – Arbeitsbericht 1:

Josef Herget

Thomas Seeger

Zum Stand der Berufsmarktforschung in der Informationswissenschaft in deutschsprachigen  
Ländern

Chur, 2007 (im Druck)

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 2

Herausgegeben von Josef Herget und Sonja Hierl

Reihe Berufsmarktforschung – Arbeitsbericht 2:

Josef Herget

Norbert Lang

Berufsmarktforschung in Archiv, Bibliothek, Dokumentation und in der Informationswirtschaft:

Methodisches Konzept

Chur, 2007 (im Druck)

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 3

Herausgegeben von Josef Herget und Sonja Hierl

Reihe Berufsmarktforschung – Arbeitsbericht 3:

Josef Herget

Norbert Lang

Gegenwärtige und zukünftige Arbeitsfelder für Informationsspezialisten in privatwirtschaftlichen  
Unternehmen und öffentlich-rechtlichen Institutionen

Chur, 2004

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 4

Herausgegeben von Josef Herget und Sonja Hierl

Sonja Hierl

Die Eignung des Einsatzes von Topic Maps für e-Learning

Vorgehensmodell und Konzeption einer e-Learning-Einheit unter Verwendung von Topic Maps

Chur, 2005

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 5

Herausgegeben von Josef Herget und Sonja Hierl

Nina Braschler

Realisierungsmöglichkeiten einer Zertifizierungsstelle für digitale Zertifikate in der Schweiz

Chur, 2005

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 6

Herausgegeben von Josef Herget und Sonja Hierl

Reihe Berufsmarktforschung – Arbeitsbericht 4:

Ivo Macek

Urs Naegeli

Postgraduiertenausbildung in der Informationswissenschaft in der Schweiz:

Konzept – Evaluation – Perspektiven

Chur, 2005

ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 7  
Herausgegeben von Josef Herget und Sonja Hierl  
Caroline Ruosch  
Die Fraktale Bibliothek:  
Diskussion und Umsetzung des Konzepts in der deutschsprachigen Schweiz.  
Chur, 2005  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 8  
Herausgegeben von Josef Herget und Sonja Hierl  
Esther Bättig  
Information Literacy an Hochschulen  
Entwicklungen in den USA, in Deutschland und der Schweiz  
Chur, 2005  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 9  
Herausgegeben von Josef Herget und Sonja Hierl  
Franziska Höfliger  
Konzept zur Schaffung einer Integrationsbibliothek in der Pestalozzi-Bibliothek Zürich  
Chur, 2005  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 10  
Herausgegeben von Josef Herget und Sonja Hierl  
Myriam Kamphues  
Geoinformationen der Schweiz im Internet:  
Beurteilung von Benutzeroberflächen und Abfrageoptionen für Endnutzer  
Chur, 2006  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 11  
Herausgegeben von Josef Herget und Sonja Hierl  
Luigi Ciullo  
Stand von Records Management in der chemisch-pharmazeutischen Branche  
Chur, 2006  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 12  
Herausgegeben von Josef Herget und Sonja Hierl  
Martin Braschler, Josef Herget, Joachim Pfister, Peter Schäuble, Markus Steinbach, Jürg Stuker  
Evaluation der Suchfunktion von Schweizer Unternehmens-Websites  
Chur, 2006  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 13  
Herausgegeben von Josef Herget und Sonja Hierl  
Adina Lieske  
Bibliotheksspezifische Marketingstrategien zur Gewinnung von Nutzergruppen:  
Die Winterthurer Bibliotheken  
Chur, 2007  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 14  
Herausgegeben von Josef Herget und Sonja Hierl  
Christina Bieber, Josef Herget  
Stand der Digitalisierung im Museumsbereich in der Schweiz  
Internationale Referenzprojekte und Handlungsempfehlungen  
Chur, 2007  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 15  
Herausgegeben von Josef Herget und Sonja Hierl  
Sabina Löhner  
Kataloganreicherung in Hochschulbibliotheken  
State of the Art Überblick und Aussichten für die Schweiz  
Chur, 2007  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 16  
Herausgegeben von Josef Herget und Sonja Hierl  
Heidi Stieger  
Fachblogs von und für BibliothekarInnen – Nutzen, Tendenzen  
Mit Fokus auf den deutschsprachigen Raum  
Chur, 2007  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 17  
Herausgegeben von Josef Herget und Sonja Hierl  
Nadja Kehl  
Aggregation und visuelle Aufbereitung von Unternehmensstrategien mithilfe von Recherche-Codes  
Chur, 2007  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 18  
Herausgegeben von Josef Herget und Sonja Hierl  
Rafaela Pichler  
Annäherung an die Bildsprache – Ontologien als Hilfsmittel für Bilderschliessung und Bildrecherche  
in Kunstbilddatenbanken  
Chur, 2007  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 19  
Herausgegeben von Josef Herget und Sonja Hierl  
Jürgen Büchel  
Identifikation von Marktnischen – Die Eignung verschiedener Informationsquellen zur Auffindung  
von Marktnischen  
Chur, 2007  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 20  
Herausgegeben von Josef Herget und Sonja Hierl  
Andreas Eisenring  
Trends im Bereich der Bibliothekssoftware  
Chur, 2007  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 21  
Herausgegeben von Josef Herget und Sonja Hierl  
Lilian Brändli  
Gesucht – gefunden? Optimierung der Informationssuche von Studierenden in wissenschaftlichen  
Bibliotheken  
Chur, 2007  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 22  
Herausgegeben von Josef Herget und Sonja Hierl  
Beatrice Bürgi  
Open Access an Schweizer Hochschulen – Ein praxisorientierter Massnahmenkatalog für  
Hochschulbibliotheken zur Planung und Errichtung von Institutional Repositories  
Chur, 2007  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 23  
Herausgegeben von Josef Herget und Sonja Hierl  
Darja Dimitrijewitsch, Cécile Schneeberger  
Optimierung der Usability des Webauftritts der Stadt- und Universitätsbibliothek Bern  
Chur, 2007  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 24  
Herausgegeben von Nadja Böller, Josef Herget und Sonja Hierl  
Brigitte Brüderlin  
Stakeholder-Beziehungen als Basis einer Angebotsoptimierung  
Chur, 2008  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 25  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Hans-Dieter Zimmermann  
Jonas Rebmann  
Web 2.0 im Tourismus, Soziale Webanwendungen im Bereich der Destinationen  
Chur, 2008  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 26  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Hans-Dieter Zimmermann  
Isabelle Walther  
Idea Stores, ein erfolgreiches Bibliothekskonzept aus England – auf für die Schweiz?  
Chur, 2008  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 27  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Hans-Dieter Zimmermann  
Scherer Auberson Kirsten  
Evaluation von Informationskompetenz: Lässt sich ein Informationskompetenzzuwachs messen?  
Eine systematische Evaluation von Messverfahren  
Chur, 2009 (im Druck)  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 28  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Hans-Dieter Zimmermann  
Nadine Wallaschek  
Datensicherung in Bibliotheksverbänden.  
Empfehlungen für die Entwicklung von Sicherheits- und Datensicherungskonzepten in  
Bibliotheksverbänden  
Chur, 2009  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 29  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Hans-Dieter Zimmermann  
Laura Tobler  
Recherchestrategien im Internet  
Systematische Vorgehensweisen bei der Suche im Internet, dargestellt anhand ausgewählter  
Fallstudien  
Chur, 2009  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 30  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Hans-Dieter Zimmermann  
Bibliotheken und Dokumentationszentren als Unternehmen:  
Antworten von Bibliotheken und Dokumentationszentren auf die Herausforderungen der digitalen  
Gesellschaft  
Chur, 2009  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 31  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Hans-Dieter Zimmermann  
Karin Garbely, Marita Kieser  
Mystery Shopping als Bewertungsmethode der Dienstleistungsqualität von wissenschaftlichen  
Bibliotheken  
Chur, 2009  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 32  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Hans-Dieter Zimmermann  
Tristan Triponez  
E-Mail Records Management  
Die Aufbewahrung von E-Mails in Schweizer Organisationen als technische, rechtliche und  
organisatorische Herausforderung  
Chur, 2009  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 33  
Herausgegeben von Robert Barth, Nadja Böller, Urs Dahinden, Sonja Hierl  
und Hans-Dieter Zimmermann  
Die Lernende Bibliothek 2009  
Aktuelle Herausforderungen für die Bibliothek und ihre Partner im Prozess des  
wissenschaftlichen Arbeitens  
Chur, 2009  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 34  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Hans-Dieter Zimmermann  
Rene Frei  
Die Informationswissenschaft aus Sicht des Radikalen Konstruktivismus  
Chur, 2009  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 35  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Hans-Dieter Zimmermann  
Lydia Bauer, Nadja Böller, Sonja Hierl  
DIAMOND Didactical Approach for Multiple Competence Development  
Chur, 2009  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 36  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Wolfgang Semar  
Michaela Spiess  
Einsatz von Competitive Intelligence in Schweizer Spitäler  
Chur, 2009  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 37  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Wolfgang Semar  
Jasmine Milz  
Informationskompetenz-Vermittlung an Deutschschweizer Fachhochschulen:  
eine quantitative Inhaltsanalyse der Curricula  
Chur, 2010  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 38  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Wolfgang Semar  
Corinne Keller  
RFID in Schweizer Bibliotheken – eine Übersicht  
Chur, 2010  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 39  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Wolfgang Semar  
Bibliotheksbau in der Schweiz 1985 – 2010  
Planung – Nutzung – Ästhetik  
Herausgegeben von Robert Barth und Iris Kuppelwieser  
Chur, 2010  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 40  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Wolfgang Semar  
Stephan Becker  
Klassifikationsraster zur Relevanzanalyse aktueller Themenanfragen an einer  
Mediendokumentationsstelle in der Schweiz  
Chur, 2010  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 41  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Wolfgang Semar  
Reihe Berufsmarktforschung – Arbeitsbericht 5:  
Iris Capatt, Urs Dahinden  
Absolventenbefragung 2010  
Bachelorstudiengang Informationswissenschaft und Diplomstudiengang Information und  
Dokumentation der HTW Chur  
Chur, 2010  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 42  
Herausgegeben von Robert Barth, Nadja Böller, Sonja Hierl und Wolfgang Semar  
Saro Adamo Pepe Fischer  
Bestandserhaltung im Film-/Videoarchiv des Schweizer Fernsehens  
Chur, 2010  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 43  
Herausgegeben von Robert Barth, Iris Capatt, Sonja Hierl und Wolfgang Semar  
Patricia Düring  
Ökonomischer Mehrwert von Bibliotheken, aufgezeigt anhand ausgewählter Dienste der Zentral-  
und Hochschulbibliothek Luzern  
Chur, 2011  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 44  
Herausgegeben von Robert Barth, Iris Capatt, Sonja Hierl und Wolfgang Semar  
Pia Baier Benninger  
Model Requirements for the Management of Electronic Records (MoReq2).  
Anleitung zur Umsetzung  
Chur, 2011  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 45  
Herausgegeben von Robert Barth, Iris Capatt, Sonja Hierl und Wolfgang Semar  
Martina Thomi  
Überblick und Bewertung von Musiksuchmaschinen  
Chur, 2011  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 46  
Herausgegeben von Robert Barth, Iris Capatt und Wolfgang Semar  
Regula Trachsler  
Angebote für Senioren in Deutschschweizer Bibliotheken  
Chur, 2011  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 47  
Herausgegeben von Robert Barth, Iris Capatt und Wolfgang Semar  
Wolfgang Semar (Hrsg.)  
Arge Alp Tagung 23.-24. September 2010, Chur  
Informationsgesellschaft und Infrastrukturpolitik im Alpenraum  
Chur, 2011  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 48  
Herausgegeben von Robert Barth, Lydia Bauer, Iris Capatt und Wolfgang Semar  
Heinz Mathys  
Jungs lesen weniger als Mädchen.  
Was können Bibliotheken gemeinsam mit den Schulen tun, um dies zu ändern?  
Chur, 2011  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 49  
Herausgegeben von Robert Barth, Lydia Bauer, Iris Capatt und Wolfgang Semar  
Anina Baumann  
Stärken und Schwächen von Discovery Diensten am Beispiel des EBSCO Discovery Service  
Chur, 2011  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 50  
Herausgegeben von Robert Barth, Lydia Bauer, Iris Capatt und Wolfgang Semar  
Reihe Berufsmarktforschung – Arbeitsbericht 6:  
Iris Capatt, Urs Dahinden  
Absolventenbefragung 2011  
Hochschule für Technik und Wirtschaft HTW Chur Weiterbildungsstudiengänge  
Informationswissenschaft.  
Externer Bericht.  
Chur, 2011  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 51  
Herausgegeben von Robert Barth, Lydia Bauer, Iris Capatt und Wolfgang Semar  
Reihe Berufsmarktforschung – Arbeitsbericht 7:  
Iris Capatt, Urs Dahinden  
Absolventenbefragung 2011  
Hochschule für Technik und Wirtschaft HTW Chur Weiterbildungsstudiengänge Management.  
Externer Bericht.  
Chur, 2011  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 52  
Herausgegeben von Robert Barth, Lydia Bauer, Iris Capatt und Wolfgang Semar  
Salome Arnold  
Auf den Spuren der Barrieren für ein barrierefreies Webdesign  
Chur, 2011  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 53  
Herausgegeben von Robert Barth, Lydia Bauer, Iris Capatt und Wolfgang Semar  
Laura Stadler  
Die Gläserne Decke in Schweizer Bibliotheken  
Chur, 2012  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 54  
Herausgegeben von Robert Barth, Lydia Bauer, Brigitte Lutz und Wolfgang Semar  
Ruth Süess  
Evaluation von Web Monitoring Tools zur softwaregestützten Informationsbeschaffung  
am Beispiel ausgewählter Open Source Web Monitoring Tools  
Chur, 2012  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 55  
Herausgegeben von Robert Barth, Lydia Bauer, Brigitte Lutz und Wolfgang Semar  
Michael Hunziker  
Approval Plans und andere Outsourcing-Formen im Bestandesaufbau an den  
Wissenschaftlichen Bibliotheken der Deutschschweiz  
Chur, 2012  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 56  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Urs Dahinden, Michael Aschwanden und Lydia Bauer  
Verpasste Chancen? Altersspezifische digitale Ungleichheiten bei der Nutzung von  
Mobilkommunikation und Internet  
Chur, 2012  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 57  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Grégoire Savary  
Eine Konservierungsstrategie für das Archiv der Siedlungsgenossenschaft Freidorf bei Muttenz.  
Eine Hilfestellung für kleine Archive mit gemischten Beständen  
Chur, 2013  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 58  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Patrick Wermelinger  
Die Georeferenzierung von Katalogdaten mit Hilfe von Linked Open Data  
Chur, 2013  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 59  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Carla Biasini  
E-Books in öffentlichen Bibliotheken der Schweiz – Determinanten der Akzeptanz bei Kunden  
Chur, 2013  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 60  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Nadja Böller  
Modell zur strategischen Analyse von Konzepten zur Förderung der Informationskompetenz durch  
Hochschulbibliotheken – MOSAIK-PRO  
Chur, 2013  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 61  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Nina Santner  
Von der Mediothek zum Recherchezentrum  
Chur, 2013  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 62  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Daniela Denzer  
Gründe für die Nichtnutzung von Bibliotheken bei Pensionierten in der Deutschschweiz  
Chur, 2013  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 63  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Verena Gerber-Menz  
Übernahme von born-digital Fotobeständen und Fotografennachlässen ins Archiv  
Chur, 2014  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 64  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Vanessa Kellenberger  
E-Shop Analytics und Erfolgsoptimierung – Die wichtigsten Kennzahlen  
Chur, 2014  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 65  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Matthias Dudli  
Open Innovation in Bibliotheken – Eine Konzeptstudie der ETH-Bibliothek Zürich  
Chur, 2014  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 66  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Sarah Carbis  
Welche Verbandszeitschrift wünschen sich die Mitglieder des BIS?  
Chur, 2014  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 67  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Yvonne Lingg  
Patientenverfügung als Informations- und Kommunikationsinstrument  
Analyse der Vielfalt sowie Dokumentation der Inhalte und Standardisierungsmöglichkeiten  
Chur, 2014  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 68  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Mara Sophie Hellstern  
Förderung von Engagement in GLAM (Galleries, Libraries, Archives and Museums) durch  
Wikipedians in Residence (WiR)  
Chur, 2014  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 69  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Philipp Trottmann  
Die epochale Trendwende: Der Benutzerrückgang an öffentlichen Bibliotheken der Deutschschweiz  
Chur, 2014  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 70  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Ursula Huber  
10 Jahre Open Access Initiative – Eine Zwischenbilanz für die Schweiz  
Chur, 2014  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 71  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Beat Mattmann  
Die Möglichkeiten von RDA bei der Erschliessung historischer Sondermaterialien  
Chur, 2014  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 72  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Diane Golay  
User-center redesign of the Biotechgate portal: a remote usability testing case study  
Chur, 2015  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 73  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Felicitas Isler  
Inklusion von Mitarbeitenden mit einer Beeinträchtigung in Bibliotheken  
Chur, 2015  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 74  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Tamara Müller  
Die Schwierigkeiten bei der Recherche im Archiv(-katalog): Ursachenforschung und  
Vorschläge zur Problembhebung  
Chur, 2015  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 75  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Benjamin Fischer  
Potential von automatischen Videoanalysen im Fussball am Beispiel der Schweizer  
Super League  
Chur, 2015  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 76  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Simon Schultze  
Videospieleturniere in öffentlichen Schweizer Bibliotheken  
Ein Pilotprojekt der St. Galler Stadtbibliothek Katharinen  
Chur, 2015  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 77  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Charlotte Frauchiger  
Barrierefreie E-Books  
Chur, 2016  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 78  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Stefanie Dietiker  
Cognitive Map einer Bibliothek  
Eine Überprüfung der Methodentauglichkeit im Bereich Bibliothekswissenschaft –  
am Beispiel der Kantonsbibliothek Graubünden  
Chur, 2016  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 79  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Sharon Alt  
Konzeption und Evaluation eines Online-Tutorial zur Förderung der  
E-Health-Literacy von Männern im Alter von 50 bis 80 Jahren  
Chur, 2016  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 80  
Herausgegeben von Wolfgang Semar und Brigitte Lutz  
Bettina Wille  
Automatisierung und Digitalisierung in den wissenschaftlichen Bibliotheken der Schweiz  
Ein Oral History Projekt  
Chur, 2016  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 81  
Herausgegeben von Wolfgang Semar  
Michael Mente  
Ansichtskarten sind Ansichtssache – Bilder, Grösse und Metadaten  
Über den Wert topografischer Ansichtskarten in Archivbeständen und  
Einsichten in Fragen ihrer archivischen Erschliessung  
Chur, 2016  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 82  
Herausgegeben von Wolfgang Semar  
Fabian Muster  
Datenstrategiemodell: Ein Referenzmodell zur Entwicklung von Datenstrategien  
Chur, 2016  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 83  
Herausgegeben von Wolfgang Semar  
Sandro Lorenzo  
Bibliotheken und Integration  
Aspekte der interkulturellen Bibliotheksarbeit und deren Einfluss auf die Integration von  
Migranten und Migrantinnen sowie Menschen mit Migrationshintergrund in der Deutschschweiz  
mit einem Fokus auf den deutschsprachigen Teil des Kantons Bern  
Chur, 2016  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 84  
Herausgegeben von Wolfgang Semar  
Johannes Reitze  
Was öffentliche Bibliotheken meinen, wenn sie vom Dritten Ort sprechen  
Chur, 2016  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 85  
Herausgegeben von Wolfgang Semar  
Simone Beeler  
Sonntagsöffnungszeiten in öffentlichen Bibliotheken in der Schweiz  
Chur, 2017  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 86  
Herausgegeben von Wolfgang Semar  
Marco Humbel  
Die Umsetzung von Open Data an Wissenschaftlichen Bibliotheken der Schweiz:  
Eine qualitative Untersuchung  
Chur, 2017  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 87  
Herausgegeben von Wolfgang Semar  
Flurina Huonder  
Medieninhaltsanalyse Big Data:  
Big Data, Datenschutz und Privatsphäre in Schweizer und US-amerikanischen Zeitungen  
Chur, 2017  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 88  
Herausgegeben von Wolfgang Semar  
Marcel Hanselmann  
Makerspaces in öffentlichen Bibliotheken:  
Eine Untersuchung der didaktischen Ziele und eine Evaluation der Technologie littleBits  
Chur, 2017  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 89  
Herausgegeben von Wolfgang Semar  
Franziska Brunner  
Überlieferungsbildung 2.0:  
Eine Untersuchung zum Mehrwert von Partizipation Dritter in staatlichen Archiven  
Chur, 2017  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 90  
Herausgegeben von Wolfgang Semar  
Marcella Haab-Grothof  
„Kleider machen BibliothekarInnen“:  
Der Einfluss von Kleidung des Bibliothekspersonals auf die Kontaktaufnahme von Benutzenden  
Chur, 2017  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 91  
Herausgegeben von Wolfgang Semar  
Sven Lenz  
Customer Engagement Analytics: Clustering User Navigation Behaviour  
Chur, 2017  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 92  
Herausgegeben von Wolfgang Semar  
Isabel Merlo  
Projektmanagement in Schweizer Bibliotheken  
Eine Untersuchung, wie Schweizer Bibliotheken Projekte managen und ein  
Projektmanagementvorschlag für die GGG Stadtbibliothek Basel  
Chur, 2018  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 93  
Herausgegeben von Wolfgang Semar  
Silvana Rüfli  
Die Usability von E-Book-Angeboten wissenschaftlicher Bibliotheken  
Eine Untersuchung am Beispiel der Universitätsbibliotheken  
St. Gallen, Bern und Zürich  
Chur, 2018  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 94  
Herausgegeben von Wolfgang Semar  
Vera Knoll  
Leichte Sprache in amtlichen Publikationen und Webseiten  
Wie ernst nehmen Verwaltungen die Leichte Sprache in der deutschsprachigen Schweiz?  
Chur, 2018  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 95  
Herausgegeben von Wolfgang Semar  
Andrea Traber  
Wie lernen studentische Bibliotheks-Nutzende und was macht für sie den optimalen  
Arbeitsplatz aus?  
Eine Studie der Lernlandschaft der Universitätsbibliothek St. Gallen  
Chur, 2018  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 96  
Herausgegeben von Wolfgang Semar  
Irina Morell  
„Für das Volk und durch das Volk?“  
Bibliotheken als Gegenstand von Volksabstimmungen und Petitionen  
Chur, 2018  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 97  
Herausgegeben von Wolfgang Semar  
Monika Rohner  
Betrachtung der Data Visualization Literacy in der angestrebten Schweizer  
Informationsgesellschaft  
Chur, 2018  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 98  
Herausgegeben von Wolfgang Semar  
Kirsten Scherer Auberson  
Counteracting Concept Drift in Natural Language Classifiers:  
Proposal for an Automated Method  
Chur, 2018  
ISSN 1660-945X

Churer Schriften zur Informationswissenschaft – Schrift 99  
Herausgegeben von Wolfgang Semar  
Hanna Kummel  
Enhancing Collaboration in Collaborative Problem-Solving with Conversational Agents  
Chur, 2019  
ISSN 1660-945X

---

## Über die Informationswissenschaft der HTW Chur

Die Informationswissenschaft ist in der Schweiz noch ein relativ junger Lehr- und Forschungsbereich. International weist diese Disziplin aber vor allem im anglo-amerikanischen Bereich eine jahrzehntelange Tradition auf. Die klassischen Bezeichnungen dort sind Information Science, Library Science oder Information Studies. Die Grundfragestellung der Informationswissenschaft liegt in der Betrachtung der Rolle und des Umgangs mit Information in allen ihren Ausprägungen und Medien sowohl in Wirtschaft und Gesellschaft. Die Informationswissenschaft wird in Chur integriert betrachtet.

Diese Sicht umfasst nicht nur die Teildisziplinen Bibliothekswissenschaft, Archivwissenschaft und Dokumentationswissenschaft. Auch neue Entwicklungen im Bereich Medienwirtschaft, Informations- und Wissensmanagement und Big Data werden gezielt aufgegriffen und im Lehr- und Forschungsprogramm berücksichtigt.

Der Studiengang Informationswissenschaft wird seit 1998 als Vollzeitstudiengang in Chur angeboten und seit 2002 als Teilzeit-Studiengang in Zürich. Seit 2010 rundet der Master of Science in Business Administration das Lehrangebot ab.

Der Arbeitsbereich Informationswissenschaft vereinigt Cluster von Forschungs-, Entwicklungs- und Dienstleistungspotenzialen in unterschiedlichen Kompetenzzentren:

- Information Management & Competitive Intelligence
- Collaborative Knowledge Management
- Information and Data Management
- Records Management
- Library Consulting
- Information Laboratory

Diese Kompetenzzentren werden im **Swiss Institute for Information Research** zusammengefasst.

## IMPRESSUM

### Verlag & Anschrift

#### Arbeitsbereich Informationswissenschaft

HTW - Hochschule für Technik und Wirtschaft  
University of Applied Sciences  
Ringstrasse 37  
CH-7000 Chur

[www.informationswissenschaft.ch](http://www.informationswissenschaft.ch)

[www.htwchur.ch](http://www.htwchur.ch)

**ISSN 1660-945X**

### Institutsleitung

Prof. Dr. Niklaus Stettler

Telefon: +41 81 286 24 61

Email: [niklaus.stettler@htwchur.ch](mailto:niklaus.stettler@htwchur.ch)

### Sekretariat

Telefon : +41 81 286 24 24

Fax : +41 81 286 24 00

Email: [clarita.decurtins@htwchur.ch](mailto:clarita.decurtins@htwchur.ch)

---