Bielefeld University

Doctoral Thesis

Integrating Socially Assistive Robots into Language Tutoring Systems

A Computational Model for Scaffolding Young Children's Foreign Language Learning

by

Thorsten Schodde

A dissertation submitted in partial fulfillment of the requirements for the degree of Doktor der Ingenieurwissenschaften (Dr.-Ing.)

at the

Faculty of Technology
Bielefeld University

December 9, 2019

Integrating Socially Assistive Robots into Language Tutoring Systems – A Computational Model for Scaffolding Young Children's Foreign Language Learning

Thorsten Schodde
Social Cognitive Systems Group
Faculty of Technology
University Bielfeld

Thesis Commitee:

- Prof. Dr.-Ing. Stefan Kopp (Bielefeld University)
- Assoc. Prof. Dr. Ginevra Castellano (University of Uppsala)
- Dr. Séverin Lemaignan (University of the West of England)
- Prof. Dr. Mario Botsch (Bielefed University)
- Dr.-Ing. Anna-Lisa Vollmer (Bielefeld University)

Date of Submission:
December 9, 2019

The paper used in this publication meets the requirements for permanence of paper for documents as specified in ISO 9706.

Integrating Socially Assistive Robots into Language Tutoring Systems

A Computational Model for Scaffolding Young Children's Foreign Language Learning

Abstract

Language education is a global and important issue nowadays, especially for young children since their later educational success build on it. But learning a language is a complex task that is known to work best in a social interaction and, thus, personalized sessions tailored to the individual knowledge and needs of each child are needed to allow for teachers to optimally support them. However, this is often costly regarding time and personnel resources, which is one reasons why research of the past decades investigated the benefits of Intelligent Tutoring Systems (ITSs). But although ITSs can help out to provide individualized one-on-one tutoring interactions, they often lack of social support.

This dissertation provides new insights on how a Socially Assistive Robot (SAR) can be employed as a part of an ITS, building a so-called ”Socially Assistive Robot Tutoring System” (SARTS), to provide social support as well as to personalize and scaffold foreign language learning for young children in the age of 4-6 years. As basis for the SARTS a novel approach called A-BKT is presented, which allows to autonomously adapt the tutoring interaction to the children's individual knowledge and needs. The corresponding evaluation studies show that the A-BKT model can significantly increase student's learning gains and maintain a higher engagement during the tutoring interaction. This is partly due to the models ability to simulate the influences of potential actions on all dimensions of the learning interaction, i.e., the children's learning progress (cognitive learning), affective state, engagement (affective learning) and believed knowledge acquisition (perceived learning). This is particularly important since all dimensions are strongly interconnected and influence each other, for example, a low engagement can cause bad learning results although the learner is already quite proficient. However, this also yields the necessity to not only focus on the learner's cognitive learning but to equally support all dimensions with appropriate scaffolding actions. Therefore an extensive literature review, observational video recordings and expert interviews were conducted to find appropriate actions applicable for a SARTS to support each learning dimension. The subsequent evaluation study confirms that the developed scaffolding techniques are able to support young children's learning process either by re-engaging them or by providing transparency to support their perception of the learning process and to reduce uncertainty. Finally, based on educated guesses derived from the previous studies, all identified strategies are integrated into the A-BKT model. The resulting model called ProTM is evaluated by simulating different learner types, which highlight its ability to autonomously adapt the tutoring interactions based on the learner's answers and provided dis-engagement cues. Summarized, this dissertation yields new insights into the field of SARTS to provide personalized foreign language learning interactions for young children, while also rising new important questions to be studied in the future.

Publications

Parts of this thesis have already been published:

- Schodde, T., Hoffmann, L., Stange, S., and Kopp, S. (2019). Adapt, explain, engage – A study on how social robots can scaffold second-language learning of children. Manuscript submitted for publication.
- de Wit, J., Schodde, T., Willemsen, B., Bergmann, K., de Haas, M., Kopp, S., Krahmer, E., and Vogt, P. (2018). The effect of a robot's gestures and adaptive tutoring on children's acquisition of second language vocabularies. *In Proceedings of the 13th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*, pages 50–58, Chicago, IL, USA.
- Schodde, T., Hoffmann, L., and Kopp, S. (2017b). How to manage affective state in child-robot tutoring interactions? *In Proceedings of the 2nd International Conference on Companion Technology (ICCT '17)*, pages 1–6, Ulm, Germany.
- Schodde, T., Bergmann, K., and Kopp, S. (2017a). Adaptive robot language tutoring based on bayesian knowledge tracing and predictive decision-making. *In Proceedings of the 12th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*, pages 128–136, Vienna, Austria.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A-BKT Adaptive Bayesian Knowledge Tracing

AR Assistive Robot

ASD Autism Spectrum Disorder

ASR automated speech recognition

ATS Affective Tutoring System

BKT Bayesian Knowledge Tracing

CBM Constraint-Based Model

cHRI Child-Human-Robot Interaction

DBDN Dynamic Bayesian Decision Network

DBN Dynamic Bayesian Network

GUI Graphical User Interface

HMI Human-Machine Interaction

HMM Hidden Markov Model

ITS Intelligent Tutoring System

KLD Kullback-Leibner divergence

POMDP Partial Observable Markov Decision Process

ProTM Probabilistic Tutoring Model for Autonomous Online Planning based on Predictive Decision-Making

RBM Rule-Based Model

RL Reinforcement Learning

SAR Socially Assistive Robot

SARTS Socially Assistive Robot Tutoring System

SIR Socially Interactive Robot

SRL Self-Regulated Learning

SVM Support Vector Machine

TTS Text-To-Speech

WOz Wizard of Oz

ZPD Zone of Proximal Development

Acknowledgments

Foremost, I would like to thank my advisor Prof. Dr.-Ing. Stefan Kopp for the continuous support in all my time as a Ph.D student, for his patience, motivation, enthusiasm, and immense knowledge. I'm very grateful for his guidance, which helped me throughout my research and writing of this thesis.

Second, I want to thank my scientific collaborators of the L2TOR project. It was a great time while working together, and I am very grateful for the fruitful discussions with insights from a variety of disciplines. They pushed me forward professionally and personally.

Third, I want to thank my fellow labmates in the Social Cognitive Systems group for the stimulating discussions, for the sleepless nights we were working together before deadlines, and for all the fun we have had in the last three and a half years.

Forth, and a bit more formal, I want to acknowledge the support by the German Research Foundation (DFG) in the Cluster of Excellence 277 "Cognitive Interaction Technology" (CITEC) at Bielefeld University, and by the EU Horizon 2020 Program in the L2TOR (www.l2tor.eu) project, grant number: 688014.

Last but not the least, I would like to thank all my friend, my parents and, in particular, my girlfriend Jacqueline for helping me to survive all the stress and not letting me give up. Thanks for all your encouragement!

Education is not preparation for life; education is life itself.

— John Dewey

1 Introduction

1.1 Motivation

Language education is a global and important issue nowadays. The globalization is proceeding and the number of people moving to another country together with their families has reached the highest level on record (cf. Stroud et al., 2018). The reasons for changing the country are various, e.g., to begin a new job or to seek asylum and find protection, but most of these people are facing the same problems. While adults can often handle their new lives and jobs by using the English language, their children will have to learn the new local language quickly to be able to use it in school. Thus, for young children it is particularly important to get appropriate language training, ideally as early as possible, since it paves the way for later academic success (cf. Esser, 2006, p. 63ff.; Hoff, 2013).

But learning a language is a complex task. It involves the learning of vocabulary as well as syntactic structures, prosodic patterns, semantic meanings and situation-dependent language use (cf. Naiman et al., 1996, p. 33ff.; p. 67). Further, language learning, like learning in general, is highly individualized and follows a trajectory that depends on, e.g., the learner's changing knowledge state, receptiveness, attention span (cf. Hooper and Umansky, 2009, Chap. 1; David Cornish et al., 2009, p. 73) or engagement (Kuh, 2003). Thus, teachers need to personalize the learning experience to be able to optimally address the individual needs of each learner. One possibility is to provide tasks neither too simple nor too hard to hold learners in the so called Zone of Proximal Development (ZPD) (cf. Vygotsky, 1978, p. 86), which is meant to optimally support their skill and knowledge development. Since working in the ZPD challenges learners with tasks slightly above their current abilities, they cannot solve these tasks on their own. To succeed within this zone the teacher has to provide temporary assistance or additional helping strategies called "scaffolding". These strategies are meant to build a supporting "scaffold" for the learner to solve the current tasks, which can be withdrawn step by step when the learner proceeds in her knowledge (Bruner, 1978; Gibbons, 2002, p. 16ff.). However, such individual support is often only possible in one-on-one interactions, but not during traditional classroom instructions.

Such one-on-one tutoring interactions guided by a domain-expert have been argued to be the most effective approach (Cohen et al., 1982; Bloom, 1984; VanLehn, 2011). In fact, students who receive individual one-on-one tutoring by a domain-expert outperform students who take part in traditional classroom instructions by two standard deviations on average (Bloom, 1984). This so-called "Bloom's two-sigma effect" is often used as a gold standard against which the effectiveness of other teaching approaches and practices are measured (cf. Hogan and Pressley, 1997, p. 109f.). However, when taking into account the standards a tutor can set for their learners, based on their background knowledge, the advantage of human one-on-one tutoring over traditional classroom instructions may be closer to 0.8 sigma (VanLehn, 2011). But still, this is a positive evidence that the undifferentiated group instructions currently performed in classrooms are exceeded by one-on-one tutoring regarding the produced learning gains. However, it is barely used in common schools. One reason is that providing individual one-on-one tutoring interactions is quite expensive in terms of time and resources, because more teachers are required to teach the same number of children as compared to traditional classroom instructions. Here, modern tutoring systems offer the potential to help out by providing personalized one-on-one learning sessions for each individual child.

These so-called Intelligent Tutoring Systems (ITSs) usually consist of a user interface, a knowledge base containing information about the learning domain and models to store information about the learner's cognitive and mental states as well as planning strategies to guide the learner through the curriculum (cf. Dede, 1986). Generally, ITSs are informed by interdisciplinary evidences from cognitive and learning science as well as computational linguistics, artificial intelligence and mathematics (cf. Graesser et al., 2012) to be able to provide personalized instructions for each individual learner as many good teachers do (cf. Conati, 2009). For this, they have to profile the child, i.e., capture the child's special needs and abilities, and to tailor the learning interaction accordingly. Indeed, previous research also highlights the importance of individual adjustments of tutoring interactions (e.g., Leyzberg et al., 2014; Gordon and Breazeal, 2015; Leyzberg et al., 2018). Even a rather simple personalization strategy of an ITS, based on a heuristic assessment of the learner's skills, was shown to yield better learning results as compared to non-personalized tutoring session (Leyzberg et al., 2014).

While most ITSs already use either simple or sophisticated approaches to personalize the tutoring interaction, they often lack social support. Traditionally, they rely on a tablet or PC to deliver the tutoring instructions and to interact with the learner (Ritter et al., 2007; Vanlehn et al., 2005). But, especially language learning is known to work best within a social interaction (cf. Rohlfing et al., 2016), which everyone experienced in their childhood while learning their mother tongue. Also in kindergartens a social interaction is used to teach either the mother tongue, a foreign or second language, e.g., while reading a picture book together. At this point a so-called Socially Assistive Robot (SAR) may be able to step in and support the learner with its social presence.

In recent years, the term SAR increasingly came up in the field of human-robot interaction. It focuses on how robots can be used to provide social, engaging, motivational and personalized long-term support to human users in various domains, such as elderly care, therapy for individuals with cognitive and/or social disorders, rehabilitation and education (Matarić, 2014). It also combines a variety of disciplines, such as robotics, medicine, social psychology, education and many more (Tapus et al., 2007) and focuses on applications in our daily lives instead of simple assembly work in factories. This evolution was made possible due to the rapid development in robotics. The machines became cheaper and more robust, while the technological development reached a point at which human-like interactions using natural language and/or nonverbal behavior get feasible now.

Consequently, when taking advantage of this technology and transfer it to an educational setting, a SAR can be used, e.g., to simulate empathy based on the learner's performance and mental state during the learning interaction (e.g., Alves-Oliveira et al., 2019). Furthermore, it can show emotions such as sadness or happiness via facial expressions (e.g., Saerbeck et al., 2010) or other modalities, e.g., colorful blinking eyes (Johnson et al., 2013), and can provide motivational support combined with appropriate task feedback (e.g., Janssen et al., 2011). Finally, it can introduce learning pauses that are filled with entertaining activities, such as joint singing, playing games or doing physical exercises together, which can help the student to recover before the next learning interaction takes place (e.g., Ramachandran et al., 2017). However, the applicability of the different social behaviors is strongly dependent on the robot's role. For example, it can act as a less knowledgeable peer within a "learning by teaching" scenario in which the learner has to explain the learning content to the robot (e.g., Jacq et al., 2016). But it can also be framed as a tutor that moderates the interaction, observes the children's learning process and provides individualized hints in a social fashion (e.g., Leyzberg et al., 2012).

In the past decade, research highlighted the potential benefits of robots applied in educational scenarios as a part of so-called Socially Assistive Robot Tutoring Systems (SARTSs) (cf. Clabaugh et al., 2015) for providing one-on-one tutoring experiences, especially for children (Belpaeme et al., 2018a). For example, the physical presence of an interactive and social robot can influence and support learning so that it becomes even more effective than learning from a classical on-screen media ITS (e.g., Han et al., 2005; Hyun et al., 2008; Kose-Bagci et al., 2009; Leyzberg et al., 2012). In fact, a robot tutor can increase learners' task performance up to 50% (Kennedy et al., 2015). But even more important is that children can learn as much from a SARTS as they can learn from an adult (Westlund et al., 2017) or a child peer (Mazzoni and Benvenuti, 2015). However, also counter evidence can be found, showing that learning with and from a SARTS is just as beneficial as learning from a tablet game (Vogt et al., 2019) or that it just yields very small learning gains (Kanda et al., 2004; Movellan et al., 2009; Gordon et al., 2016). These contradictory results can be caused by different study setups, necessary limitations and the environment during the interaction, e.g., noisy classrooms at school (cf. Gordon et al., 2016; Vogt et al., 2019). But especially when embedding a SAR into a learning interaction some limitations are always indispensable. For example, adding too many social behaviors to the robot can distract the learner and harm the learning process in general (Kennedy et al., 2015). But still, using the right experimental design and strategies to make use of the robot's physical and social presence can boost learning. This provides a first indicator that learning from SARs is qualitatively different compared to other digital tutoring technologies. This is partly due to children's tendency to treat robots as peers in long-term interactions (Tanaka and Matsuzoe, 2012a) and to connect with them on a friend-like basis (Kanda et al., 2004), which, in particular, underlines its potential usefulness for language learning.

The idea of extending a traditional ITS with a SAR and using its presence to create a motivating and interactive language learning experience was already picked up in several studies (e.g., Hyun et al., 2008; Alemi et al., 2014; Westlund et al., 2015; Jacq et al., 2016), but most of them address an older target group of children. Consequently, it is an open question whether the gained knowledge from these studies is also applicable for kindergarten children, since the development of young children's cognitive and mental abilities is continuously and fast moving (cf. Mooney, 2013, Chap. 4). The same applies to insights gained from other educations fields, in which a broad range of studies have already investigated how SARs can be used to support learning (e.g., Serholt et al., 2013; Chen et al., 2018; Ramachandran et al., 2018; Senft et al., 2018). Despite their promising results, the used strategies, e.g., to encourage the student to think aloud during problem solving (Ramachandran et al., 2018), can often not be applied to foreign or second language learning or might not show the same beneficial effects.

The remaining insights and findings just result in a small body of knowledge of how a SAR can be applied for young kindergarten children within this particular field, which is not sufficient yet to provide an optimal tutoring experience. This is, inter alia, due to the multidimensionality of the "learning" process in general (cf. Kennedy et al., 2016). It does not only include the "cognitive learning" that describes the learner's knowledge gain (Bloom et al., 1956; Krathwohl, 2002; Anderson et al., 2001) but also the "affective learning" that consists of the learner's long-term motivation and interest in the learning task (Kratwohl et al., 1964; Gorham, 1988), and the "perceived learning" that describes the learner's self-perception of skill and knowledge changes during learning (cf. Alavi et al., 2002; Davidson et al., 2003, p. 243ff.). The major problem is that all three dimensions as well as the learner's engagement are strongly correlated and each teaching or scaffolding action, either performed by a robot or an ITS, probably influences more than just one dimension.

In summary, one key problem is the lack of systematic knowledge about how an ITS can be extended by a SAR to scaffold and ensure young children's language learning in all of these facets, especially given that suboptimal robot behaviors that are not matching or too distracting for the child can even hamper learning (Kennedy et al., 2015).

1.2 Research Goals

The present thesis, written in the context of the L2TOR project[1] aims for developing a novel approach to extend a traditional ITS with a SAR for scaffolding kindergarten children's foreign language learning. This includes to find out how a robot-supported tutoring interaction should be designed as well as which actions a robot and ITS can use to scaffold and foster foreign language learning in all its dimensions. To that end, a computational model that serves as the basis for a SARTS is required, which allows for an autonomous interaction by continuously adapting to the learner. More precisely, this thesis aims for investigating the following research questions:

[1] http://www.l2tor.eu

RQ0: What are relevant elements of language learning practices in German kindergartens and how can they be implemented into a SARTS to provide a meaningful basis for foreign language learning interactions?

RQ1: How can a SARTS optimally address the cognitive learning of kindergarten children?

- **RQ1.1:** How can a SARTS keep track of the individual knowledge and needs of kindergarten children to build up a sophisticated student model?
- **RQ1.2:** How can a SARTS be enabled to select appropriate actions to adapt the interaction based on the student model?

RQ2: Which scaffolding strategies can be used by a SARTS to optimally support affective and perceived learning as well as the engagement of kindergarten children?

- **RQ2.1:** What are relevant affective and cognitive states, as well as behavioral cues, to track the engagement of kindergarten children during foreign language learning interactions?
- **RQ2.2:** Which actions can be used by a SARTS to scaffold the engagement and affective learning of kindergarten children during foreign language learning interactions?
- **RQ2.3:** Which strategies can be applied by a SARTS to scaffold the perceived learning of kindergarten children during foreign language learning interactions?

RQ3: How to combine multiple strategies addressing the different learning dimensions into a single approach that is capable of modeling the learning dimensions' interconnections and allows for an autonomous interaction with online adaptation?

To address these questions several studies were conducted starting with a data collection to define an appropriate language learning scenario for kindergarten children as a basis for all following studies. Since the major goal of a tutoring interaction is to teach knowledge, a novel and easily extendable model is presented and evaluated, which can serve as a basis for an ITS or SARTS, respectively. It allows to track children's knowledge state and, based on this, to autonomously make decisions to plan and adapt the next steps of the tutoring interaction. In addition, this model allows to represent the influences of tutoring and scaffolding strategies on the learning process as well as between the different learning dimensions so that they can also be considered during the planning process. However, since there is still a lack of systematic knowledge about which tutoring and scaffolding actions work best for kindergarten children and how they should be designed and implemented, especially when applied by a robot, new strategies had to be designed and evaluated. They are meant to positively influence the learner's engagement, affective and perceived learning, which in turn can foster cognitive learning.

1.3 Contributions

This thesis contributes to the field of ITSs and provides further insights on how such systems can be enriched with a SAR to support children's learning. While this field already contains a huge amount of research, the informational body of systematic knowledge on how a SAR can be applied to the foreign language learning of kindergarten children is still limited. Here, this thesis starts to fill up this knowledge gap through the following novel contributions:

- **Tutoring interaction design for SARTSs:** to be able to investigate the different research questions, a suitable structure and setting for the tutoring interaction as well as guidelines to provide appropriate feedback to kindergarten children are required. Furthermore, this information needs to be transferable to a SARTS. Therefore, this thesis presents an empirical basis derived from observational recordings in German kindergartens, which includes the different aspects that need to be considered when implementing a suitable tutoring interaction for young kindergarten children.

- **Autonomous and adaptive teaching model:** the central part of an ITS contains separate modules to keep track of the learner's developing knowledge and to plan the next steps of the remaining tutoring interaction. This thesis presents a novel approach to combine both, knowledge tracing and decision-making, into a single and easily extendable model. This yields the possibility to simulate the effects of all teaching actions based on the system's belief about the learner's current knowledge and to optimally plan the next steps. The results of two evaluation studies show that a personalized tutoring interaction guided by this model is able to significantly improve students' learning performance and to maintain their engagement on a significantly higher level as compared to the control conditions with non-adaptive tutoring.

- **Scaffolding strategies for Socially Assistive Robots:** Since learning is a multidimensional process, a tutoring system has not only to address the student's cognitive learning but also to support the remaining dimensions, e.g., by providing additional scaffolding. Consequently, this thesis presents new insights gained from domain experts on how the engagement of kindergarten children can be tracked, which behavioral cues are important and which scaffolding strategies, e.g., re-engaging the child, can be used by a SARTS to react appropriately. In addition, a novel scaffolding strategy to address students' perceived learning is developed and implemented. It is based on literature of transparency and open learner models, and is meant to support learners' self-perception as well as to reduce their uncertainty. An evaluation study with kindergarten children demonstrated that the implemented scaffolding strategies can indeed be used to either maintain or recover learners' engagement. Further, they allow to support the perceived learning of slow learning children and, with that, their learning process so that they can keep up with their stronger classmates.

- **Fusion of the adaptive teaching model and scaffolding strategies:** The multidimensionality of the learning process makes it indispensable for an autonomously acting ITS to consider the different influences of the learning dimensions, as well as of all included scaffolding and tutoring actions. Thus, this thesis presents an extended version of the developed autonomous and adaptive teaching model that integrates and also allows to adapt to these influences. More precisely, applying this model enables an ITS to simulate the effects of different action combinations, e.g., first re-engage the child before providing the next teaching task, and to choose the most profitable action series to optimally foster learning while maintaining an appropriate level of engagement. The model's behavior was tested via simulations of different learner types, which were derived from studies with kindergarten children. The corresponding results nicely illustrate the model's ability to identify the individual needs of each learner type and to tailor the interaction accordingly.

1.4 Thesis Outline

This thesis is organized as follows. In Chapter 2 the theoretical background, with respect to the different learning dimensions, the engagement, their interconnections and the important concepts of language learning, is introduced. Chapter 3 introduces Intelligent Tutoring Systems, their possible structures and fields of application. Furthermore, it defines the term of Socially Assistive Robots, the related work on Socially Assistive Robot Tutoring Systems and their application in educational settings. Subsequently, Chapter 4 focuses on the empirical basis that contributes to the development of a suitable interaction structure, tutoring game and feedback behaviors for kindergarten children. In Chapter 5 the related work and development of a novel approach for knowledge tracing and predictive decision-making are presented. Furthermore, in this chapter two user studies are presented, one conducted with adults and one with kindergarten children, which evaluate the developed model. The following Chapter 6 describes the related work, development and evaluation of scaffolding strategies applicable by a Socially Assistive Robot to support young children in their engagement, affective and perceived learning. Afterwards, Chapter 7 proposes an extended version of the model presented in Chapter 5, which integrates all findings of the present thesis. This model is evaluated with simulations of different learners derived from the previous studies and the corresponding results are presented and discussed. Furthermore, current limitations and required future investigations to fit the missing parameters of this model are discussed. Finally, Chapter 8 summarizes this thesis, highlights its contributions as well as its limitations and points towards future research directions.

Since this thesis was written within the scope of the L2TOR project, several colleagues were involved in the work described in the Chapters 4 to 6. Thus, the author's contributions will be noted in the following. First of all, the author contributed to the analysis of pre-recordings conducted in German Kindergartens presented in Chapter 4. Furthermore, he provided strong and independent contributions in a small team to the subsequent conception of the tutoring design, as well as the development of the general system setup. In Chapter 5, the author was fully responsible for the development and implementation of the described A-BKT model, as well as the study system for the first evaluation study (see Section 5.3) and did almost all of this work himself and independently. Furthermore, he provided strong and independent contributions in a small team to all stages of the study, namely, conception, planning and conduction of the actual study, as well as the analysis and interpretation of the results. The author's implemented study system was, with small adaptations, also used in the second evaluation study (see Section 5.4), in which he further provided small independent contributions to the planning process, as well as the analysis and interpretation of the respective results during joint work with a project partner from the Netherlands. In addition, the author also contributed to the analysis and interpretation of the subsequent engagement analysis study (see Section 5.5). In the remaining two studies presented in Chapter 6, namely, the expert interviews (see Section 6.2) and the scaffolding evaluation study (see Section 6.3), the author was again fully responsible for the development, as well as the implementation of the study systems and did almost all of this work himself and independently. Finally, he provided strong and independent contributions in a small team to all stages of both studies including the planning, the study conduction, as well as the analysis and interpretation of the results.

Education is the kindling of a flame, not the filling of a vessel.

— Socrates

2 Background

Developing a SARTS to support kindergarten children in their language learning requires a sufficient understanding of how learning works in general. The following chapter summarizes this information, which provides insights into the potential needs of the learner and, thus, is indispensable when building a SARTS that creates an optimal and personalized tutoring experience for language learning. This includes knowledge about the different dimensions of learning (cognitive, affective and perceived learning) and the learner's engagement, how they are interconnected and how they can be addressed (Section 2.1). Additionally, more specific knowledge about important concepts of language and word learning is required, which also adds further requirements to the learning interaction (Section 2.2).

2.1 Dimensions of Learning and Engagement

Following general findings on learning, it can be assumed that learning with SARTS also involves the three different dimensions of learning, namely, cognitive, affective and perceived learning (cf. Kennedy et al., 2017a). While cognitive learning typically refers to the knowledge and skills addressed in learning interactions (Bloom et al., 1956; Anderson et al., 2001; Krathwohl, 2002), affective learning represents aspects, such as attitudes, values and motivation towards the learning objective or content (Kratwohl et al., 1964; Gorham, 1988). Perceived learning, instead, focuses on the learner's beliefs of how much she has learned and how confident she is about her learned knowledge (cf. Alavi et al., 2002; Davidson et al., 2003, p. 243ff; Caspi and Blau, 2008, p. 327). Sometimes the third dimension is replaced by the psychomotor domain that includes, inter alia, physical movements and coordination tasks (cf. Dave, 1970; Harrow, 1972; Simpson, 1972), which can range from easy tasks, such as washing a car, to more complex tasks, such as dancing. However, since language education is not directly connected to the psychomotor domain, this thesis focuses only on perceived learning instead.

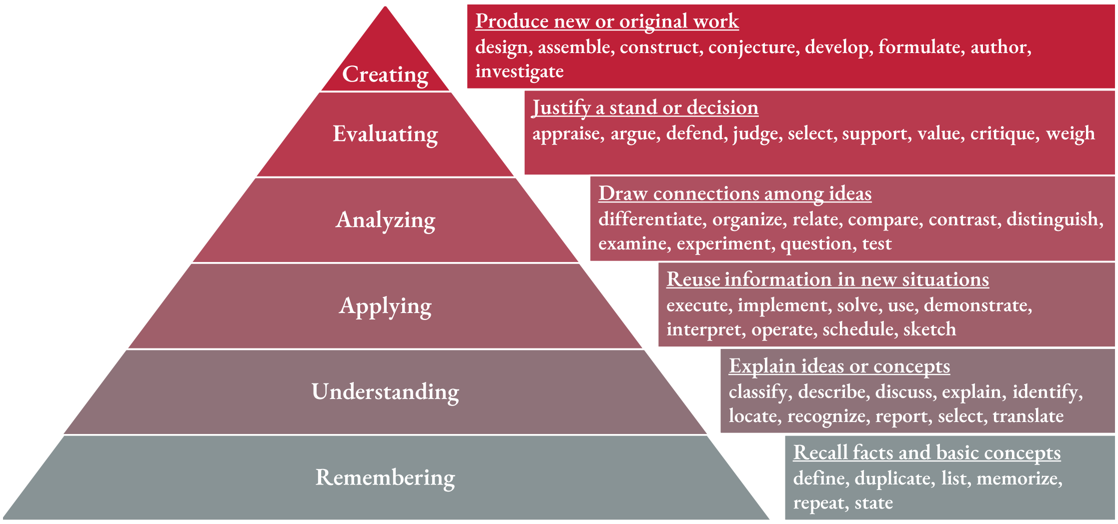

Creating
Evaluating
Analyzing
Applying
Understanding
Remembering
Produce new or original work
design, assemble, construct, conjecture, develop, formulate, author, investigate
Justify a stand or decision
appraise, argue, defend, judge, select, support, value, critique, weigh
Draw connections among ideas
differentiate, organize, relate, compare, contrast, distinguish, examine, experiment, question, test
Reuse information in new situations
execute, implement, solve, use, demonstrate, interpret, operate, schedule, sketch
Explain ideas or concepts
classify, describe, discuss, explain, identify, locate, recognize, report, select, translate
Recall facts and basic concepts
define, duplicate, list, memorize, repeat, state

Figure 2.1: The six revised learning categories from Anderson et al. (2001) based on Bloom's taxonomy from easy (gray) to more complex (red). The rectangles on the right provide further explanations and verbs associated with the respective categories (taken and redesigned from de Vicente (2018)).

2.1.1 Cognitive Learning

According to Bloom et al. (1956) the concept of cognitive learning involves the to be learned knowledge and the development of intellectual skills. This includes recognizing or recalling facts, procedures, patterns and concepts that help to develop intellectual abilities. To describe this process, Bloom et al. (1956) developed the "Taxonomy of educational objects", which covers the following six major categories of cognitive learning: (1) the *knowledge*, which "involves the recall of specifics and universals, the recall of methods and processes, or the recall of a pattern, structure, or setting" (Bloom et al., 1956, p. 201); (2) the *comprehension*, which "refers to a type of understanding or apprehension such that the individual knows what is being communicated and can make use of the material or idea being communicated without necessarily relating it to other material or seeing its fullest implications" (Bloom et al., 1956, p. 204); (3) the *application* that refers to the "use of abstractions in particular and concrete situations" (Bloom et al., 1956, p. 205); (4) the *analysis*, which represents the "breakdown of a communication into its constituent elements or parts such that the relative hierarchy of ideas is made clear and/or the relations between ideas expressed are made explicit" (Bloom et al., 1956, p. 205); (5) the *synthesis* of knowledge that involves the "putting together of elements and parts so as to form a whole" (Bloom et al., 1956, p. 205); (6) finally, the *evaluation*, which evokes "judgments about the value of material and methods for given purposes" (Bloom et al., 1956, p. 207). These categories can be thought of as degrees of difficulties one has to achieve step by step, from concrete to abstract levels in a cumulative hierarchy so that the previous level(s) must be mastered before the next one can be tackled.

A few decades later, Bloom's established theory has been revised by Anderson et al. (2001). In the traditional taxonomy the categories are only based on the knowledge while the corresponding underlying cognitive processes were only mentioned as supplements in the hierarchy. To break up this static structure, Anderson et al. (2001) separated the knowledge and the cognitive processes and created two single dimensions for them. They further renamed and restructured the old categories, and allow for the new hierarchy to be more flexible so that categories may overlap. Moreover, the new six categories focus on the underlying cognitive processes of cognitive learning and, therefore, are renamed with gerunds (see Figure 2.1). In detail, they can be described as follows: (1) *remembering* the acquired knowledge and being able to recognize and recall it later on; (2) *understanding* the knowledge so that it can be used by the learner to interpret, exemplify, classify, summarize, infer, compare and explain it; (3) *applying* the newly learned knowledge to execute or implement something in a new situation; (4) *analyzing* new situations to differentiate them from or organize and attribute them to already acquired knowledge; (5) *evaluating* materials by means of checking them for correctness and criticizing them; (6) planing and *creating* new content based on the attained knowledge (Anderson et al., 2001, p. 14ff.). As can be seen, *synthesis* was renamed to *creating* and changed place with *evaluation*. This is because *creating* involves inductive thinking, which is a more complex cognitive task than deduction that is usually used when one *evaluates* something.

In contrast, the second dimension represents the knowledge applied/learned in each cognitive process and can be divided into the following four types: (A) *factual knowledge* is defined as knowledge about terminologies, specific details and elements of the learning content; (B) *conceptual knowledge*, which includes knowledge of classifications and categories, principles and generalizations, as well as theories, models and structures with respect to the learning domain; (C) *procedural knowledge* consists of subject-specific skills, algorithms, techniques and methods, as well as criteria for determining when to use appropriate procedures; (D) *metacognitive knowledge*, which summarizes strategic knowledge, self-knowledge and knowledge about cognitive tasks, including appropriate contextual and conditional knowledge (Anderson et al., 2001, p. 27ff.).

With the resulting two dimensional representation it is possible to create a so-called taxonomy table with knowledge as vertical and the six cognitive processes as horizontal dimension. Based on this table the difficulty level of different tasks can be analyzed by categorizing their required knowledge and cognitive processes. For instance, given the exemplary task taken from Krathwohl (2002)

> "Write original compositions that analyze patterns and relationships of ideas, topics, or themes."
>
> (Krathwohl, 2002, p. 216)

the following knowledge (gray) and cognitive processes (red) can be identified (see Table 2.1). In the noun phrase the required knowledge is described, whereas the verb phrases contain the needed cognitive processes to solve the given task. In detail, within the noun phrase, the words "patterns and relationships" belong to the *conceptional knowledge* and can be sorted in accordingly. The verb "analyze" belongs to the forth cognitive process *analyzing*, whereas the verb "write" belongs to *creating*. The resulting taxonomy table is presented in Table 2.1.

This taxonomy table helps to identify all cognitive processes and types of knowledge addressed by a learning task and allows for creating a matching tutoring interaction in which the SARTS can track the learner's mastery of needed skills and adapt the interaction accordingly, e.g., by choosing appropriate tasks and/or providing additional support.

	remembering	*understanding*	*applying*	*analyzing*	*evaluating*	*creating*
factual knowledge						
conceptual knowledge				X		X
procedural knowledge						
metacognitive knowledge						

Table 2.1: The resulting taxonomy table for the example given by (Krathwohl, 2002, p. 216). The addressed *conceptual knowledge* has to be *analyzed* and used to *create* a composition.

2.1.2 Affective Learning

The dimension of affective learning is also indispensable for education and covers parts of the learner's emotional and belief system. Practitioners such as Kratwohl et al. (1964), for instance, describe the affective domain as "objectives which emphasize a feeling tone, an emotion, or degree of acceptance or rejection [...] expressed as interests, attitudes, appreciations, values, and emotional sets or biases" (Kratwohl et al., 1964, p. 7). Although these types of objectives are often difficult to measure in quantifiable terms, many educators try to cope with them, because they want their students to "appreciate" what they are learning and to "feel good" about themselves during class. In fact, most educators recognize during their own empirical teaching practice that learning occurs more often and to a greater extent when their students are emotionally involved, which is also supported by neurobiological studies (cf. Levy, 1983; Davidson and Cacioppo, 1992, see also Section 2.1.4). Hence, it is important to address the affective learning dimension during teaching and it can be assumed that a lack of emotive stimuli may cause the learner to give up their efforts for sustained learning.

The affective learning domain is also represented in the taxonomy of Bloom et al. (1956) and Kratwohl et al. (1964). It is explained as a paradigm, which can be divided into five major categories (see Figure 2.2), again ordered from simple to complex: (1) *receiving phenomena*, which means that the learner is aware of the learning interaction, has the willingness to hear about the learning content and focuses her attention to the learning problem; (2) *responding to phenomena* is describing the active participation of the learner, so that she is attending and reacting to particular phenomena arising during the learning interaction. Furthermore, a good learning outcome may improve the willingness to respond and the satisfaction in responding (motivation); (3) *valuing* describes the value a learner attaches to the learning process, which ranges from simple acceptance to a more complex state of commitment. Although it is based on the learner's internal set of particular values, clues to theses values are often expressed by the learner through overt behaviors and, thus, are identifiable by teachers; (4) *organization* refers to assigning priorities to values by contrasting them, resolving conflicts between them, and creating an unique value system. An example for this is the recognition of an imbalance between free time and other responsibilities. The learner then creates a life plan in harmony with abilities, interests and beliefs to prioritize the time effectively to meet the needs of work, family and self; (5) *internalizes*

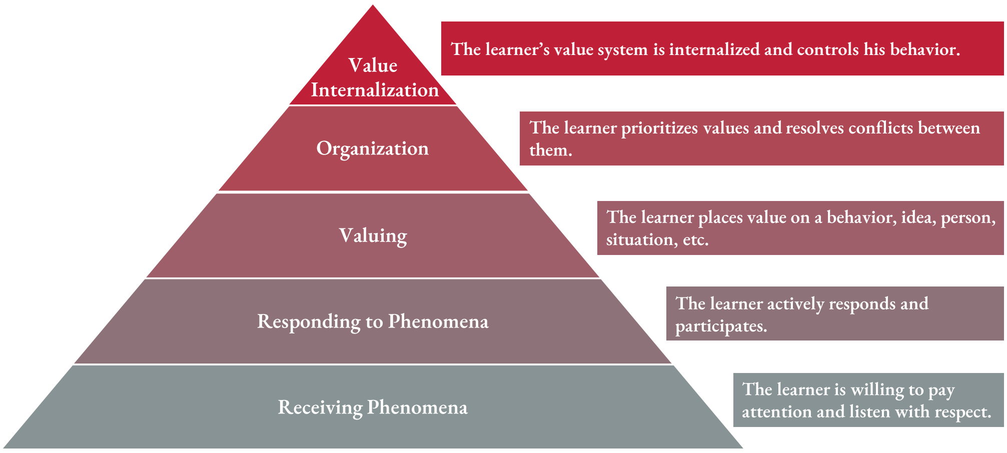

Value Internalization
The learner's value system is internalized and controls his behavior.
Organization
The learner prioritizes values and resolves conflicts between them.
Valuing
The learner places value on a behavior, idea, person, situation, etc.
Responding to Phenomena
The learner actively responds and participates.
Receiving Phenomena
The learner is willing to pay attention and listen with respect.

Figure 2.2: The five categories contained in the affective learning dimension (Kratwohl et al., 1964) ordered from easy (gray) to more complex (red). Further explanations for the different categories are presented in the rectangles on the right.

values (characterization) so that the learner has an internal value system influencing her behaviors to be pervasive, consistent and predictable. Additionally, instructional objectives focus on learners' general patterns of adjustment on the personal, social and emotional level. Examples for this category are that the learner shows self-reliance when working independently, shows teamwork abilities while working in a group, uses an objective approach in problem solving or shows professional commitment to ethical practice on a daily basis (Kratwohl et al., 1964, p. 34f.).

Although the taxonomies for the cognitive and affective learning describe different, apparently independent categories or objectives, they are closely interconnected and influence each other. Additionally, separating both becomes already somewhat artificial, because no teacher or curriculum worker intents one entirely without the other when preparing the learning content. However, examining each dimension separately yields a good impression of what aspects have to be included in a good teaching interaction. For example, it can be used by a SARTS to adapt the difficulty of the learning content or to provide appropriate scaffolding for the cognitive and affective learning dimension (see Section 2.1.5). This in turn can lead to a feeling of "flow" (cf. Basawapatna et al., 2013), which can further increase the learner's positive value towards learning and rise her motivation to increase her endeavor.

2.1.3 Perceived Learning

The dimension of perceived learning, also known as *perceived ability* (e.g., Miller et al., 1996; Greene and Miller, 1996), refers to a retrospective evaluation of the learning experience and can be defined as a "set of beliefs and feelings one has regarding the learning that has occurred" (Caspi and Blau, 2008, p. 327). It is often described in connection with the concept of Self-Regulated Learning (SRL), in which students try to adapt their own learning behavior, e.g., the application of solving strategies or deciding on the next learning content. In general, this is called the *self-reflection phase*, which includes two different aspects (Davidson et al., 2003, p. 243ff). First, *self-reactions* that mostly consist of emotions, e.g., feeling ashamed due to the impression of putting too little effort into a task while not being able to solve it successfully and, second, *self-evaluations*, which include mainly objective measures of the reached learning outcome. The latter refers to self-monitoring one's outcomes and comparing them with a goal or standard, which has to be achieved. While this is relatively easy in tasks where only one correct outcome exists, e.g., in mathematics or vocabulary learning, this might be harder in more complex tasks or tasks including a social component. In general, four major criteria for *self-evaluations* can be differentiated: (1) the *mastery*, which is an absolute measure of the solution's quality, e.g., comparing a solved math problem with the intended solution. If the measurement of the absolute mastery is not possible, e.g., because of an unstructured informal context, (2) the *previous performance* can be used for *self-evaluation* by comparing it with current performance. This is also called *self-criteria* (cf. Ellis and Zimmerman, 2001, p. 212) and requires a performance record of the past learning endeavors, e.g., assessing the growing competence in solving crossword puzzles by counting the remaining gabs and errors made by comparing it to the author's solution. (3) The *normative criteria* is defined as social comparison of one's solution with solutions of others, such as classmates or in competitions, in which one gets awarded, e.g., with a medal at a spelling bee. (4) The *collaborative criterion* is primary used while working in teams (Bandura, 1991). Here, the success is defined in terms of the fulfillment of a particular role in the team. For example, in an Academic Olympic competition where every person is an expert of a different topic so that the criterion of problem-solving success might be different for a "science expert" compared to a "humanities expert" (Davidson et al., 2003, p. 244f).

These *self-evaluations* or even *self-judgments* are also related to the *causal attributions* of the effort's outcomes, meaning, whether a failure is due to one's limited ability or to insufficient endeavors. This is crucial for the learning interaction, because attributing errors to a fixed learning ability might cause students to react negatively, discourage them and let them give up (cf. Weiner, 1979). However, attributing errors to the applied solving strategy has been shown to be effective in sustaining motivation during problem solving (e.g., Zimmerman and Kitsantas, 1996, 1997), because it sustains a perception of efficacy until all possible strategies are tested. But in general, the *attribution* is not automatically a result of favorable or unfavorable *self-evaluations*. It further depends on the affective learning dimension, since prior motivational beliefs can also affect the outcome expectations (Bandura, 1991).

The last important aspect in self-perception and *attribution* is the *intentionality* (cf. Weiner, 1985). If students, for instance, do not intent to solve a task properly, although they have the skill to do so, they probably will not experience the adverse *self-reactions* for a weak self-perception resulting from a poor performance. Whereas students, who fail at a problem while intending to solve it, might attribute this to a lack of ability and, thus, might experience shame and dissatisfaction (Davidson et al., 2003, p. 245). Consequently, the concept of intentionality is able to influence or modify the general effects of self-perception and *attribution* on the learner's emotions and affective learning dimension.

In Summary, the perceived learning has a strong influence on the overall learning process and, thus, can be used by a SARTS to further support the learner to achieve her goals. For example, it can support the learner in understanding her own knowledge and motivate her so that she is learning with the right mindset and attribute her errors correctly. This in turn can lead to a more positive attitude towards learning, higher endeavor to solve the tasks and, with that, to higher learning gains.

2.1.4 Engagement

In addition to general motivational aspects, also learning related affective states, such as boredom or frustration, have to be taken into account, since they also influence the learning interaction significantly (Craig et al., 2004). They are often summarized under the concept of engagement, which is allied with the affective learning dimension. In recent years, this concept has frequently been used in a variety of settings, such as human-human (Glas and Pelachaud, 2014), human-agent (Hall et al., 2005; Peters et al., 2005) and also human-robot interactions (Sanghvi et al., 2011; Le Maitre and Chetouani, 2013). However, the variety of definitions is as broad as the areas of application (see Glas and Pelachaud (2015) for an overview). Goffman (1966), for instance, defined engagement in terms of face engagement:

> "Face engagements comprise all those instances of two or more participants in a situation joining each other openly in maintaining a single focus of cognitive and visual attention – what is sensed as a single mutual activity, entailing preferential communication rights."
>
> (Goffman, 1966, p. 89)

Even though this definition of engagement is relatively old, it is still applied nowadays (e.g., Couture-Beil et al., 2010; Le Maitre and Chetouani, 2013), especially in interactions where people employ eye-contact, gaze and facial expressions to interact with each other (Le Maitre and Chetouani, 2013). Also other definitions can be found that can be applied on a meta-level, e.g., by including the attribution of a value onto the interaction goal (Peters et al., 2005):

> "[...] the value that a participant in an interaction attributes to the goal of being together with the other participant(s) and of continuing the interaction."
>
> (Peters et al., 2005, p. 1)

This definition is frequently applied in systems and studies (e.g., Bohus and Horvitz, 2009; Glas and Pelachaud, 2014) or serves as a basis for engagement tracking opportunities (e.g., Castellano et al., 2009; Sanghvi et al., 2011).

However, according to educational psychologists, such as Linnenbrink and Pintrich (2003), engagement can generally be divided into three major areas: (1) the *behavioral engagement*, which involves the learner's observable behavior with respect to "Are they working hard? Are they distracted? Do they show help seeking behavior? Do they endure a task even if they encounter difficulties or do they give up easily?". Usually teachers can easily tell if students are behaviorally engaged by simply watching them and they try to maximize it, because students who are showing higher behavioral engagement persist longer at tasks and reach higher learning gains (Skinner and Belmont, 1993). (2) The *cognitive engagement*, instead, describes, for instance, whether the students are not only behavioral engaged and show attention, e.g., by gazing at the teacher, but also think deeply about the learning content and try to use different strategies to increase their learning gains and understanding of the materials. (3) Finally, the *motivational engagement* focuses on learners intrinsic motivation to learn within a particular setting, which can be achieved by choosing appropriate content and tasks in terms of the learner's interest, skill level and affect (Linnenbrink and Pintrich, 2003).

When comparing the definition provided by Linnenbrink and Pintrich (2003) and the first three concepts of the affective learning domain multiple similarities can be found. However, in many cases the engagement does not only represent a subset of the affective learning dimension but also includes a broader set of emotional aspects influencing the learning interaction as stated by Hall et al. (2005):

> "Empathic engagement is the fostering of emotional involvement intending to create a coherent cognitive and emotional experience which results in empathic relations between a user and a synthetic character."
>
> (Hall et al., 2005, p. 1f.)

This definition already includes deeper emotional states and refers to the cognitive and emotional experiences the user senses. Especially these emotional experiences are important in general learning interactions. Studies showed that emotional or affective states, such as confusion, anxiety, flow, boredom, frustration and surprise are able to influence student's answering behavior and learning gains (e.g., Craig et al., 2004; Lehman et al., 2010). Further, such states influence cognitive processes like attention, understanding, remembering, long-term memorization, reasoning and the application of knowledge during task solving (Tyng et al., 2017). Consequently, students' emotional and affective states should definitely be considered carefully when developing a SARTS, since they strongly influence the learning process. For example, a SARTS can work pro-actively, e.g., by adjusting the task difficulty to establish a feeling of flow so that it is not too easy nor too hard, which in turn can prevent boredom and frustration (cf. Csikszentmihalyi, 2014, p. 243ff.). Further, the SAR can be used to express emotions, which can result in an overall higher positive valance and, thus, in positive affective states (Gordon et al., 2016).

In summary, learners' affective states play a crucial role within learning interactions. Although they are not directly responsible for high or low learning gains, they strongly influence the learning process as a whole. In general, it can be concluded that a high engagement with positive affective states can be highly beneficial for a learning interaction and, thus, the definition of engagement in the scope of foreign language tutoring has to take learners' affective states into account and, with that, goes beyond the affective learning dimensions. However, since the definition of engagement is also dependent on further information, e.g., about the tutoring setting and the robot's role, the definition used in this thesis is defined in Chapter 6 after all required information are collected.

2.1.5 Scaffolding

As briefly mentioned, to provide a beneficial learning interaction for each individual learner appropriate scaffolding is required that allows to address the engagement and the different learning dimensions. In this context, the term scaffolding refers to any strategy or temporary assistance provided by a teacher to guide and support the learner in moving towards new skills, concepts or levels of understanding (Bruner, 1978; Gibbons, 2002, p. 16ff.). More specifically, scaffolding means to build a "scaffold" for

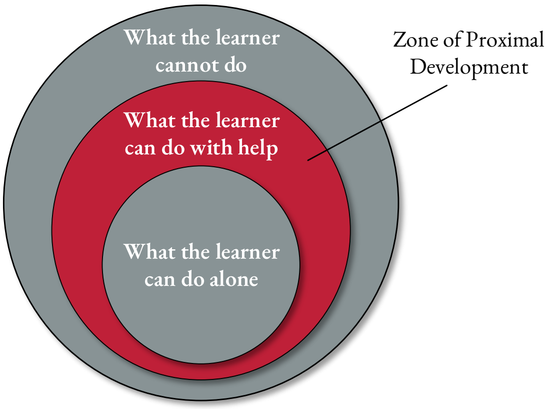

What the learner cannot do
What the learner can do with help
What the learner can do alone
Zone of Proximal Development

Figure 2.3: Illustration of the ZPD. Students can either learn on their own or together with a tutor in the ZPD to solve tasks slightly above their current abilities or can be faced with tasks they cannot even solve with external help (taken and redesign from Culatta (2011)).

the learner by means of supportive actions, such as giving examples, demonstrating or modeling task solutions, highlighting the important aspects of a task, or breaking a task down into simpler sub-tasks for the learner (Wood et al., 1976). It further includes to provide hints and partial solutions to encourage the learner to think-aloud or to give direct instructions how to solve a task (Hartmann, 2002, Chap. 3). Gibbons (2002) even argues that "it is only when teacher support – or scaffolding – is needed that learning will take place, since the learner is then likely to be working within his or her zone of proximal development" (Gibbons, 2002, p. 16f.). Vygotsky (1978) defined this particular zone as "the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem-solving under adult guidance, or in collaboration with more capable peers" (Vygotsky, 1978, p. 86, see also Figure 2.3). Thus, scaffolding performed by human tutors and probably also a tool such as a SARTS can support students to work in the ZPD and to learn new skills to an extend beyond what they could have achieved without this help. If scaffolding was successful and the learner has internalized the new skill, the scaffold can be withdrawn or adapted to lead towards new skills that should be acquired next.

The previously mentioned examples for scaffolding mainly address the dimension of cognitive learning. However, this concept is not limited to this domain and can also be applied to influence the affective and perceived learning. For example, to provide scaffolding for students' affective learning and engagement, a teacher can try to motivate the learner before providing a task by sketching the learning goal and the benefits gained when achieving it. She further can monitor and control the learner's frustration during problem solving, e.g., by providing additional help or scaffolding at the right time (cf. Bransford et al., 2000, pp. 104). In addition, the teacher can provide appropriate feedback for the learner, which can be defined as "information provided by an agent regarding aspects of one's performance or understanding" (Hattie and Timperley, 2007, p. 81). In general, feedback can be provided in two different ways (cf. Pat-El et al., 2013). First, it can just be informative by reflecting the "monitored" performance of the learner relative to the learning goals, i.e., "where you are" and "where to go" (Sadler, 2009). Second, the feedback can be enhanced by additional scaffolding to support the learner by giving direction and advice (Shepard, 2005), i.e., “how to get there” (Sadler, 2009). Both have been shown to positively affect the learner’s intrinsic motivation (Shute, 2008; Corbalan et al., 2009), while the information of how to improve in a task achieved the highest positive effect (Moreno, 2004; Dresel and Haugwitz, 2008).

However, providing information about the learner’s current state of knowledge and already attained goals can also be used to address the dimension of perceived learning. According to Lin et al. (1999), this information does not necessarily need to be provided through verbal feedback from a single teacher and can also be communicated via process displays or a forum for reflective social discourse, e.g., a classroom. Furthermore, they argue that this type of additional scaffolding for *self-reflection* or perceived learning, respectively, can support students during learning and improve their development of adaptive learning expertise by increasing their reflective practice (Lin et al., 1999). Moreover, raising the student’s intrinsic motivation by providing appropriate feedback can also serve as a scaffold for learners’ *intentionality* (see Section 2.1.3), since it influences or modifies the general effects of self-perception and *attribution of errors* on the affective learning dimension and engagement. Consequently, it might be beneficial either to motivate students verbally before or during the learning interaction or to use an easy task on which they will succeed, so that they are working with the right mindset and motivation to interpret their self-perception and *attribute their errors* correctly.

2.1.6 Correlations between Learning Dimensions and Engagement

The presented studies and definitions already provided hints that strong correlations between the learning dimensions and the engagement exist (see Figure 2.4 for an overview). To prevent undesirable influences of applied tutoring and scaffolding actions, which might lower the effectiveness of a learning interaction, this important aspect has to be considered, too. In particular, the cognitive and affective learning dimensions are strongly correlated with the student’s engagement and highly influence each other (Schwarz, 2000; Craig et al., 2004; Lehman et al., 2010; Hamari et al., 2016). For instance, Craig et al. (2004) found a significant relationship between cognitive learning and the affective states of confusion, flow and boredom. While boredom can hamper learning, a low amount of confusion can support cognitive learning, which in turn can result in a good learning progress and a feeling of flow (Craig et al., 2004). This in turn can result in a positive mood and a higher intrinsic motivation (cf. Csikszentmihalyi, 2014, p. 233ff.; p. 255), which can cause a positive value and attitude towards learning (see Section 2.1.2) and, thus, can support the cognitive learning dimension.

But both dimensions together with the engagement influence and are also influenced by the learner’s self-perception (perceived learning). On the one hand, it was shown that a good self-perception ability during learning is a strong predictor for school achievements and, thus, for cognitive learning (Pajares and Miller, 1994; Spinath et al., 2006). However, negative observations during learning can also lower the learner’s engagement (Miller et al., 1996; Greene and Miller, 1996), which in turn can cause her to attribute the observed problem to her limited skills, to loose the learning motivation and to let her stop the endeavors (cf. Davidson et al., 2003). On the other hand, motivational prior beliefs and engagement can influence the self-perception either positively or negatively (Bandura, 1991; Schwarz, 2000). If stu-

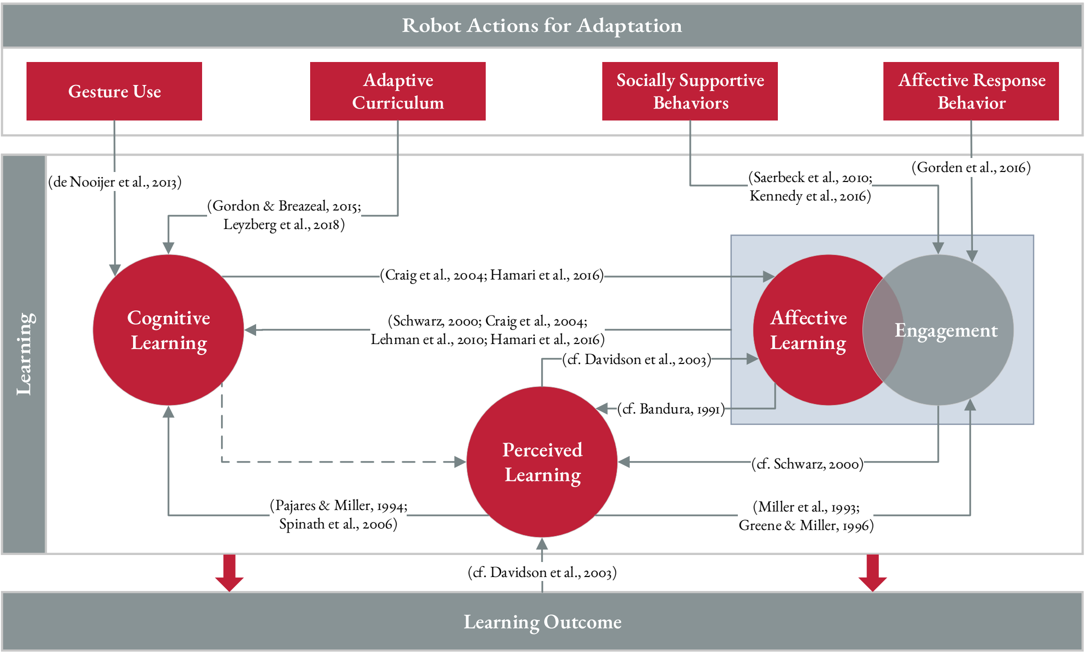

Robot Actions for Adaptation
Gesture Use
Adaptive Curriculum
Socially Supportive Behaviors
Affective Response Behavior
Learning
(de Nooijer et al., 2013)
(Gordon & Breazeal, 2015; Leyzberg et al., 2018)
(Saerbeck et al., 2010; Kennedy et al., 2016)
(Gorden et al., 2016)
(Craig et al., 2004; Hamari et al., 2016)
Cognitive Learning
(Schwarz, 2000; Craig et al., 2004; Lehman et al., 2010; Hamari et al., 2016)
(cf. Davidson et al., 2003)
Affective Learning
Engagement
(cf. Bandura, 1991)
Perceived Learning
(cf. Schwarz, 2000)
(Pajares & Miller, 1994; Spinath et al., 2006)
(Miller et al., 1993; Greene & Miller, 1996)
(cf. Davidson et al., 2003)
Learning Outcome

Figure 2.4: The different dimensions of learning and their interrelations. The arrows visualize reported influences between dimensions of learning and engagement, as well as effects of a SARTS's actions on all of them. Since the learner's engagement partially shares aspects with the affective learning, they are depicted by overlapping each other (taken and redesigned with permission from Schodde et al. (2019)).

dents, for instance, have a negative attitude towards learning and show low engagement, they probably also interpret the observed outcomes more negatively resulting in a worse self-perception compared to students, who have a positive attitude towards learning so that it is more valuable for them.

Hence, to reduce the negative and strengthen the positive effects, a SARTS needs actions to influence and manage the different learning dimensions. Several studies already investigated such actions and their effects (cf. Figure 2.4), e.g., the use of gestures (de Nooijer et al., 2013), adaptation of the curriculum (Gordon and Breazeal, 2015; Leyzberg et al., 2018), socially supportive behaviors (Saerbeck et al., 2010; Kennedy et al., 2016) or affective feedback behavior (Gordon et al., 2016). But nevertheless, the body of knowledge about which actions of a SAR are beneficial, which influences do these actions have on the learning dimensions and which side effects combinations of different actions might cause is still limited and needs to be extended to optimally take advantage of a SAR included in an ITS.

But, the learner's engagement, the learning dimensions and their interconnections are not the only aspects to be considered. Each learning field has its own interaction dynamics and concepts that also require careful consideration when designing a SARTS.

2.2 Language Learning

Using speech for communication purposes is an ability, which has been develop quite late in the human history. It is a result of an interplay of a variety of cognitive processes and has to be learned or acquired, because it is not a skill anchored in our genes. However, it is an inherent endeavor of each human to communicate by using either a language or non-verbal behaviors such as facial expressions and gestures. Even newborns are already interested in social communications to express their feelings and needs and can already recognize faces, touches and also sound (cf. Müller, 2013, p. 44f.).

In general, three different types of languages have to be differentiated. First, the mother tongue, which is acquired in a mainly unconscious process in the first years of life. Second, the foreign language, which is learned in school, but is not present in the learner's daily life so that its access is limited. Finally, the second language that is both, acquired and learned, which happens, e.g., when parents speak different languages and decide to raise their children bilingual (cf. Oxford, 2003, p. 1). Especially the last type demonstrates that both processes, acquiring and learning, are not fully distinct and can take place at the same time (cf. Krashen, 1981, p. 1ff.). Further, they preferably take place within a social interaction, which is one of the most important aspects of language learning in general. Roseberry et al. (2009), for instance, showed that children younger than 3 years only learn the meaning of verbs in a close interaction with their mother. This is, when presenting the verbs within a video, even though still spoken by their mothers, they did not understand their meaning correctly (Roseberry et al., 2009). Older children, however, are already able to understand the meaning of verbs through a video, although they still showed higher learning gains in direct interactions (Roseberry et al., 2009; cf. Müller, 2013, p. 53).

2.2.1 Learning a Foreign Language

In general, young children are able to learn new languages faster than adults and achieve a learning speed similar to the acquisition of their mother tongue. Furthermore, they can achieve high competencies in the use of grammar and complete phoneme sets (pronunciation) (Ghasemi and Hashemi, 2011; Müller, 2013, p. 65f.). This observation is supported by the theory about the existence of an *critical phase* for language learning. Lenneberg (1967) described this phase as a limited timespan in which it becomes possible to learn a further language to an extend comparable to the mother tongue (Lenneberg, 1967). However, the age-region of this critical phase is controversial discussed. While most researchers agree on having this phase located in childhood, the start and end points range from the age of three to four up to the puberty (cf. Kim, 2007, p. 5f.;Müller, 2013, p. 65f.).

A newer theory called *neural commitment* further supports the existence of this phase and describes the development of neural patterns, which mainly depends on the perception frequency of multiple languages. These patterns can influence the ability to learn languages later on, meaning, if children mainly hear their mother tongue their brains mostly provide specialized systems to filter particularly these familiar phonemes and, thus, do not allow for an easy generalization to other languages (cf. Kuhl, 2004). This is also supported by other studies, which provide further evidence that learning an additional language quite early can significantly influence the structure of our speech systems (cf. Mechelli et al., 2004; Schlegel et al., 2012; Mayer et al., 2015).

But, although these studies provide first hints, it is still not fully resolved how learning of further languages in different stages of children's development can influence the organization and representation of the mother tongue, a foreign or a second language in their brains. Müller (2013), for instance, presented three possible constellations. All languages might either share the same area, just a subset or are even located in fully distinct areas (cf. Müller, 2013, p. 58ff.). However, first evidence is provided that adults who have a high proficiency in both languages use a shared brain area, while those with a lower proficiency in the foreign language use separate cerebral areas for the representations (cf. Kim, 2007, p. 6). This further supports the *neural commitment* theory in the sense that parts of the speech system can generalize and handle multiple languages, while this still does not provide any information about the optimal age for learning further languages.

In the past, critics even feared that learning or acquiring two or more languages in parallel might cognitively overstrain children so that they are not able to master even one language completely. However, the mastery of a language is hard to measure, since it is not well defined yet. Even the language skills of native speakers can vary although they did not learn a second or foreign language. In addition, a lot of studies addressed and refuted the critics' apprehension and, instead, highlighted the benefits. They have shown that the simultaneous exposure to two languages increases the phonological awareness, fosters children's thinking about language per se and leads to higher meta-cognitive and meta-linguistic skills (Bialystok, 2007; Ramirez and Kuhl, 2016). In addition, the majority of the world population is raised bilingual and no significant verbal deficits can be found[1]. Hence, the benefits of early language education outweighs the apprehension of critics and it seems to be reasonable to start early with learning a foreign or second language. This cannot only lead to higher proficiency, in the ideal case comparable to the mother tongue, but also to more generalizable neural patterns and representations in their brains, which help to learn and speak new languages and enable the children to be optimally prepared for their later lives.

2.2.2 Word Learning

A major aspect of language education is the learning of new words, their syntax and semantics. While it is still not fully resolved why especially young children are able to learn their mother tongue and its vocabulary so quickly (Kuhl, 2004), a variety of theories have been developed to partially demystify this process. Capone and McGregor (2005), for instance, introduced the concept of *fast mapping*:

> "Fast mapping is the initial association of word and referent in memory. Fast mapped (and infrequently encountered) words are incompletely represented with limited semantic and lexical knowledge and few connections to other words in memory."
>
> (Capone and McGregor, 2005, p. 1469)

The major point of this theory is that in most cases a single contact to a word is already enough to create a first rough concept of this particular word in our brains. Especially, young children are quite good in using these rough concepts, because they are trained to learn a lot of words each day. In particular, they are able to associate words with them already after a few contacts (cf. Horst and Samuelson, 2008). After a first rough concept is established, the process called *slow mapping* is applied

[1] Deutsche Gesellschaft für Sprachwissenschaft (2018): "Mehrsprachigkeit" – Können Kinder nur eine Sprache gleichzeitig learnen? URL: https://dgfs.de/de/thema/bilingualer-erwerb.html

to enrich it with more details, e.g., about different contexts of application (cf. Singleton, 2012, p. 279f.; Lüke and Ritterfeld, 2014, p. 203f.; Müller, 2013, p. 51).

Another concept for word learning is developed by Rohlfing et al. (2016). The so-called *pragmatic frame* describes word learning as an embedded process in a specific activity or setting:

> "A pragmatic frame is a negotiated interaction protocol targeted to achieve a joint goal [...]"
>
> (Rohlfing et al., 2016, p. 2)

This theory further highlights the importance of a social interaction for word learning in the sense that using a specific word frequently within such an interaction can additionally support its memorization. To establish a pragmatic frame it is important to define a mutual interaction goal to allow for the learner to analyze and understand each single step and activity taken in which the target words can be embedded. This leads to a familiarization effect so that the interaction structure is easily interpretable, which results in lower cognitive load for the learner and, hence, allow her to concentrate on the word learning itself (Rohlfing et al., 2016).

Although both concepts describe different aspects, they are often applied simultaneously, e.g., while "reading" a picture book. The adult (parent or educator) asks the child about names of certain objects or if not known yet, gives them names, while the child has to remember them quickly. Here, the underlying interaction concept of reading any kind of picture book serves as the *pragmatic frame* with which most children are rapidly familiarized. Additionally, the child has to use the process of *fast mapping* to combine the frequently occurring new concepts with newly given names into rough concepts, which are enriched with further details later on (*slow mapping*).

2.3 Summary

Investigating the effects and influences of all three dimensions of learning (cognitive, affective and perceived) highlights the necessity to take all into account when building a SARTS. Although the intended progress during tutoring primary takes place in the cognitive learning dimension, the others can either support or hamper this progress. Moreover, the student's engagement needs to be considered as well. Although it overlaps with the affective learning dimension, it goes beyond it and considers the learner's affective states, which also influences the learning process significantly. Consequently, to create an optimal language learning interaction a tutoring system needs to provide multidimensional support, e.g., by applying appropriate scaffolding actions. This further allows for the learner to work in the ZPD and enables her to achieve goals beyond what she could have achieved without this support.

However, selecting appropriate teaching and scaffolding actions is a complicated task, since all learning dimensions are strongly interconnected. That is, each tutoring or scaffolding action executed by a SARTS can influence multiple dimensions, which in turn complicates the planning process. For instance, motivating the learner will lead to higher affective learning, which can increase cognitive learning. Providing good explanations and scaffolding during teaching can foster cognitive learning, which can lead to a more positive attitude towards the subject, enhance the motivation to interact with the

SARTS (affective learning) and, additionally, foster positive affective states (engagement). This in turn can result in a higher learning gain and a feeling of performing well (perceived learning). Conversely, actions that cause a negative attitude towards the subject (affective learning) can lower engagement, which can further hamper cognitive and then perceived learning. Consequently, the strong interconnections can not only result in great synergy effects but also in unintended negative influences so that it is indispensable to consider these correlations carefully when building a SARTS or an ITS in general.

In addition, general findings on language learning should be considered as well. Although research supports that it is reasonable to encourage already young children to learn a foreign or even second language, since it is much easier for them as compared to adults and yields a positive influence onto their neural speech system, they still can be further supported by a SARTS. For example, the tutoring system can use the SAR to provide a social presence for establishing a *pragmatic frame*, which allows for the learner to benefit from a familiarization effect so that she can fully concentrate on the learning content instead of thinking about the general interaction. Further, the concept of *fast mapping* can be used so that the learner has a first rough idea about the learning content, before the actual tutoring interaction starts, in which the content is repeated several times to trigger the *slow mapping* process.

In conclusion, the identified information about the different learning dimensions, learners' engagement, as well as the knowledge about language and word learning provides a meaningful basis for the development of a SARTS. First of all, the information serve as a guideline regarding important aspects of learning and can be used as a basis for the development of different adaptation capacities to tailor the tutoring interaction to the individual needs of each learner. However, before a SARTS can be developed further information about the general structure of ITSs and the possible paradigms to implement them is required. In addition, knowledge is needed that includes the already investigated benefits of SARs, how they were used to enrich tutoring settings and which adaptation capacities of SARTSs with respect to the different learning dimensions were already studied. This is summarized in the following chapter.

We're fascinated with robots because they are reflections of ourselves.

— Ken Goldberg

3 Related Work

The background information with respect to the three dimensions of learning, the engagement, their strong interconnections and the important concepts of language learning summarized in the previous chapter can be regarded as guidelines to follow throughout the development of a SARTS. However, developing a SARTS that can provide multidimensional support further requires information about the general structure of ITSs and how SARs can be applied to establish a social interaction. Therefore, this chapter focuses first on the general formalisms of ITSs to get an impression of the different implementation paradigms and the required modules to build such a system (Section 3.1). Afterwards, the definition of SARs is discussed and their important abilities are highlighted. Moreover, a review of the corresponding literature is presented to identify their effects (benefits/disadvantages) in a variety of application fields (Section 3.2). This further includes literature about the benefits and challenges for the utilization of SARs to construct a SARTS that can be applied in educational settings, and, in particular, in language learning interactions for children (Section 3.3). Finally, the advantages of different adaptation and personalization abilities of these systems are highlighted and discussed with respect to the different dimensions of learning (Section 3.3.2).

3.1 Intelligent Tutoring Systems

An Intelligent Tutoring System (ITS) aims for providing personalized tutoring instructions and feedback for each individual learner (cf. Psotka et al., 1988). Classical definitions for ITSs are provided by Conati (2009):

> "Intelligent Tutoring Systems (ITS) is the interdisciplinary field that investigates how to devise educational systems that provide instruction tailored to the needs of individual learners, as many good teachers do. Research in this field has successfully delivered techniques and systems that provide adaptive support for student problem solving

> in a variety of domains. There are, however, other educational activities that can benefit from individualized computer-based support, such as studying examples, exploring interactive simulations and playing educational games. Providing individualized support for these activities poses unique challenges, because it requires an ITS that can model and adapt to student behaviors, skills and mental states often not as structured and well-defined as those involved in traditional problem solving."
>
> (Conati, 2009, p. 1)

and also by Graesser et al. (2012):

> "Intelligent Tutoring Systems (ITS) are computerized learning environments that incorporate computational models in the cognitive sciences, learning sciences, computational linguistics, artificial intelligence, mathematics, and other fields. An ITS tracks the psychological states of learners in fine detail, a process called student modeling."
>
> (Graesser et al., 2012, p. 2)

In general, an ITS can be seen as "the initiative to apply artificial intelligence to education and instructional design" (Paviotti et al., 2012, p. 17). Thus, its major tasks are to reason about all the information provided by the learner, to consult the curriculum and, based on this, to decide what should be done next. A variety of ITSs were developed in the past decades, such as the Andes system (Vanlehn et al., 2005), which supports learners during their physics homework, Cognitive Tutor (Ritter et al., 2007), which provides adaptive scaffolding on algebra problems, or Wayang (Arroyo et al., 2004), which offers a huge number of homework questions from mathematics including matching feedback and randomized control tests for practice purposes (see also Anderson et al. (1995) for an overview).

3.1.1 Architectures of Intelligent Tutoring Systems

Although the definitions of ITSs describe their general tasks, they lack information about their internal structure. Traditionally, one of the following two different architectures is used. The first architecture consists of three blocks or components, namely, the systems domain expertise, tutoring expertise and a module that stores information about the student's knowledge and skill (cf. Derry et al., 1988; Siemer and Angelides, 1998). Derry et al. (1988), for instance, used this architecture and proposed an ITS that includes an *expert domain model*, which contains guidelines from experts, and a *student knowledge model* that provides information about the learner's knowledge state. Both are used to inform a *tutoring model*, which plans an individual path through the curriculum and selects the actions to be used next (Derry et al., 1988). Although their modules are named differently, parallels of the modules' tasks can still be found between the initially defined three-model architecture and that of Derry et al. (1988).

The second and more common architecture for building ITSs simply extends the previously mentioned three-module architecture by adding a *user interface* as a fourth component (cf. Dede, 1986; Nwana, 1990; Freedman et al., 2000; Nkambou et al., 2010, see Figure 3.1). It was first proposed by Dede (1986) and allows to take the learner's input into account and to provide all necessary task information. The remaining modules include the *knowledge base*, also known as the *domain model*,

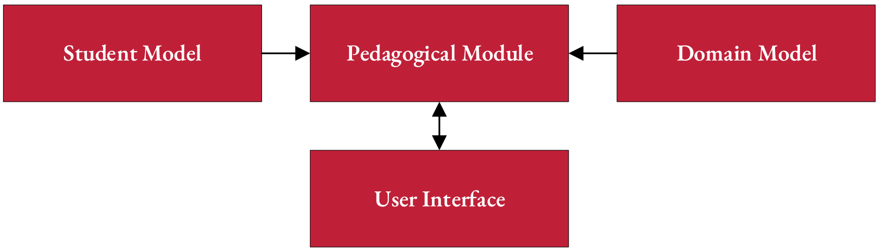


Figure 3.1: Four module architecture for ITSs from Dede (1986). The *pedagogical module* considers the information stored in the *domain* and *student model* to plan the next steps of the tutoring interaction. Further, the ITS uses *user interface* to communicate with the learner.

which provides the declarative (what), procedural (how) and meta-cognitive (thinking about what and how) knowledge. Further, a *student* or *cognitive model* is included that stores information about the learner's knowledge and comprehension, her cognitive processes (calculation and problem solving), meta-cognitive strategies like learning from errors and/or psychological attributes (development level, learning style, emotions). And, finally, the *pedagogical* or *tutoring module* that makes use of the information about the learner to find an efficient path through the curriculum. Additionally, this module employs tutoring strategies based on the learner's evolving knowledge state and an underlying instructional theory that allows to determine the effects of each pedagogical tutoring strategy (cf. Dede, 1986).

3.1.2 Paradigms for Implementing Intelligent Tutoring Systems

In general, the literature on ITSs distinguishes between three major approaches for representing and reasoning about the learning domain (cf. Nkambou et al., 2010). The first approach, called Rule-Based Models (RBMs), consists of a set of rules the student has to follow step by step while solving a problem (e.g., Anderson et al., 1995; Vanlehn et al., 2005; Crowley and Medvedeva, 2006; Koedinger and Aleven, 2007; Aleven, 2010). These models are usually used if the learning problem can be structured into well defined and easily verifiable steps (Nkambou et al., 2010, p. 33ff.). An example for this are Model-Tracing Tutors, which are also based on a set of production rules, but further tries to follow the learner's reasoning process by analyzing the applied rules during problem solving. This allows to represent the learner's reasoning or even cognitive processes in depth (e.g., Ritter et al., 2007). Moreover, Model-Tracing Tutors can support a wide variety of tutoring scenarios and can enhance the tutoring interaction by (1) evaluating the learner's knowledge in terms of applied skills, (2) inferring the learner's goals, (3) suggesting the next step to be taken and (4) giving demonstrations to the learner (Nkambou et al., 2010, p. 85). Model-Tracing Tutors are mainly applied when the goal is to evaluate the reasoning process rather than simply determining whether the learner already attained the taught knowledge. However, one of the major limitations of this approach is that it becomes very complex and time consuming to specify a sufficient set of rules and solution paths in huge domains.

In contrast, the second paradigm called Constraint-Based Model (CBM) only defines a set of requirements that all solutions have to satisfy (e.g., Mitrovic, 2003; Suraweera et al., 2005; Mitrovic et al., 2007; D'Mello et al., 2008). This simplifies the implementation process significantly, since it is possible to represent the whole space of correct answers by a few constraints. In addition, CBMs split up the solution space so that all final states, in which the tutor has to execute the same actions (e.g., give feedback), are grouped together in the same class. But this results in a crucial assumption: the actual sequence of actions taken by the learner is not important for being able to find mistakes. Instead, it is enough to only observe the final state of the given solution (Nkambou et al., 2010, p. 63ff.). Thus, compared to RBMs that often accept only one specific way of solving a problem as the correct answer, CBMs provide greater freedom. That is, each constraint that needs to be satisfied by the learner's answer in order to be correct only focuses on a small part of the learning domain and, hence, it is much easier to define and check constraints in CBMs compared to RBMs. But this benefit also results in losing a lot of details about the learner's proficiency when using CBMs.

Finally, the third approach integrates an expert system into an ITS (e.g., Clancey, 1984; Graesser et al., 2000; Moritz and Blank, 2008; Ghadirli and Rastgarpour, 2013) that emulates the ability of a human expert's decision-making by simulating their operative modalities, as well as their skills in modeling and facing problems. In general, two different use cases for employing expert systems can be found in the literature. (1) The expert system is used to generate several solution-paths, which the ITS compares with the learner's answers. Further, the generated solutions can be used as demonstrations or suggestions to support the learner during the problem solving process or to plan the next steps of the interaction. An example for this is the GUIDON system, which is used to teach the expertise from a knowledge base consisting of more than 500 rules given by medical practitioners to students in an organized, efficient and comprehensive way (Clancey, 1984). A second example is provided by Ghadirli and Rastgarpour (2013), who proposed an E-Learning ITS, which employs an expert system that selects the learning content and techniques with respect to the learner's knowledge and preferred learning style (Ghadirli and Rastgarpour, 2013). (2) The expert system is utilized to compare just the ideal solution with the learner's answer and, with this, to provide detailed feedback about possible problems. This approach was applied, e.g., in the DesignFirst-ITS (Moritz and Blank, 2008) and Auto-Tutor (Graesser et al., 2000). But although expert systems yield a lot of benefits, they also have their limitations. For instance, "(1) developing or adapting an expert system can be costly and difficult, especially for ill-defined domains; and (2) some expert systems cannot justify their inferences, or provide explanations that are appropriate for learning" (Nkambou et al., 2010, p. 88).

3.2 Socially Assistive Robots

Socially Assistive Robots (SARs) provide the potential to be applied as a social output component for traditional ITSs. They can be seen as a combination of two related fields intersecting each other (see Figure 3.2). First, Assistive Robots (ARs) that can, for instance, help humans by lifting objects in their private homes or factory environments, or support them in physical rehabilitation, e.g., in form of a wheelchair robot (Simpson and Levine, 1997, cf. Feil-Seifer and Mataric, 2005). Hence, their focus is more on physically assisting humans. In contrast, the goal of so-called Social Robots or Socially Interactive Robots (SIRs) is to establish and maintain a good social interaction with people. They can

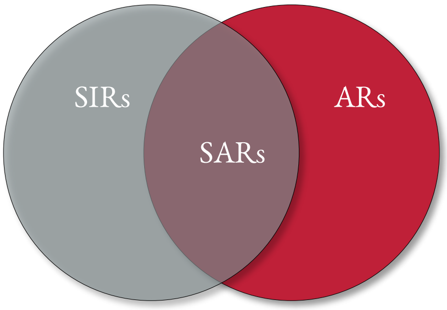


Figure 3.2: The field of Socially Assistive Robots (SARs) can be seen the intersection of Socially Interactive Robots (SIRs) and Assistive Robots (ARs).

take several roles and can act as peers or partners with various shapes (cf. Fong et al., 2003; Feil-Seifer and Mataric, 2005, see also Section 3.2.1).

The field of SARs combines parts of both, establishing and maintaining a social interaction, while assisting people in various life or job settings. The latter is also the key difference to SIR. While the goal of SIR is the social interaction itself, SARs aim for establishing a social interaction to provide assistance and to achieve measurable progress. Furthermore, instead of assisting the user physically, which is the goal of ARs, SARs focus more on assisting through their social presence and socially supportive behaviors (cf. Feil-Seifer and Mataric, 2005). For example, SARs can provide social and personalized interaction, e.g., by offering motivational and engaging long-term support for the user, to assist in various tasks, such as elderly care, rehabilitation or learning (cf. Feil-Seifer and Mataric, 2005; Tapus et al., 2007; Fasola and Matarić, 2013; Matarić, 2014; Clabaugh et al., 2015; Gordon and Breazeal, 2015). However, this combination also results in the necessity to address the challenges provided by both fields. On the one hand, SARs have to face the situation-dependent challenges of each task in which they should provide assistance. On the other hand, they have to handle the new challenges arising from interacting with social individuals, such as human beings.

3.2.1 Challenges in Social Robotics

The terms of Social Robotics, Social Robots or Socially Interactive Robot have become synonyms for an upcoming field of research in the past decades. Robots are not only meant to do assembly work in factories anymore but also start to arrive in our daily lives. But this new field of application also rises new challenges, initiated through the necessity to interact with social beings. For example, they have to recognize and understand the user while acting in line with social rules and norms. This is also reflected in the definitions of Social Robots, which focus not only on the robot's task and role but also on the challenges and requirements they have to fulfill. For instance, Social Robots can be defined as follows:

> "A social robot is an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact."
>
> (Bartneck and Forlizzi, 2004, p. 592)

This definition already highlights the new abilities a social robot has to have. First, it needs to be able to interact and communicate with humans, preferably in a human like and natural way or at least in a fashion that is understandable by humans. Thus, it needs modalities to produce verbal output and/or non-verbal behaviors, such as nodding or pointing. Second, while doing so, it has to choose its actions based on behavioral norms and rules accepted by humans, which are quite more complex compared to the set of rules applied in a controlled factory environment. Finally, it has to behave semi-autonomously or even completely autonomously.

Del Moral et al. (2009) extended this definition by also adding the environment and physical constraints of the robot:

> "A Social Robot is an autonomous motion device equipped with sensors, actuators and interfaces (robot) that interacts and communicates with humans following some expected behavior rules, which are founded on the robot physical properties and the environment within it is embedded, mainly taking into account the needs of the people with witch it is meant to interact"
>
> (del Moral et al., 2009, p. 5)

Thus, if a social robot is also mobile, e.g., by moving on wheels or even on its own legs, it needs further sensors to be aware of its environment and also extended navigation skills to be able to fulfill its tasks without interfering or harming its interaction partners. Additionally, it needs appropriate actuators and interfaces to be able to interact with its environment and also with humans.

Although the navigation through social environments and the interaction with humans yield a lot of new challenges, the application of social robots also provides a lot of benefits. In contrast to an interaction with a tablet screen or a PC, the interaction with a social robot feels more natural to humans. A robot, for instance, can use common non-verbal cues, such as eye-gaze to establish joint attention, nod to show agreement or other types of gestures. However, one might argue that these behaviors can also be used by virtual agents, but research has shown that robots are perceived more helpful, credible, informative and enjoyable to interact with (Kidd and Breazeal, 2004; Wainer et al., 2007).

3.2.2 Fields of Application

Although a social interaction increases the difficulty of developing social robots and, thus, also SARs, they were already applied in a variety of settings. For example, in health or convalescent care they can provide social support by distracting or engaging patients during their recovering periods in hospitals (Saldien et al., 2006) or support and teach them to handle their diseases (Henkemans et al., 2013; Broadbent et al., 2018). The ALIZ-E project[1], for instance, used the Nao robot (see Figure 3.3a) to provide personalized health education for children with diabetes (Henkemans et al., 2013). Although non-social robots have also been shown to yield good results in rehabilitation (Lo et al., 2010) or health care (Davies, 2016), there might be still a big potential to improve their effectiveness with a social component by making the therapeutic process more enjoyable (cf. Matarić et al., 2007).

[1] http://www.aliz-e.org/

(a) Nao[2]

[redacted]

[redacted]

[redacted]

[redacted]

Figure 3.3: Socially Assistive Robots used in health (a,b) and elderly care (c,d,e).

Another field of application is the elderly care in which SARs can be used to entertain people or to help them with their mental health issues, while trying to maintain their independence as long as possible (cf. Libin and Cohen-Mansfield, 2004; Feil-Seifer and Mataric, 2005; Banks et al., 2008; Lehmann et al., 2013; Broadbent et al., 2014; Jenkins and Draper, 2014; Chang and Šabanović, 2015; Orejana et al., 2015). However, the elderly are often distrustful and, thus, hesitate or even decline to interact with a robot (cf. Miehle et al., 2019). In these cases the robot's social abilities can be used to lower their hesitation and to raise their acceptance of the robot's role and assistance, which, in fact, was already verified in several studies and even in long-term interactions (e.g., Broadbent et al., 2014; Orejana et al., 2015). But this requires to design the robots and their behaviors carefully to allow for a significant increase in the quality of life or medication adherence in this particular setting (cf. Broadbent et al., 2014). However, despite this impediment several robots with various shapes were already applied to this setting. They are ranging from large machine like robots, such as NurseBot Pearl (see Figure ??) that is used to relieve nurses in their day to day activities by reminding the elderly to take their medication and guiding them around the environment (Montemerlo et al., 2002; Pineau et al., 2003), down to small pet-like robots, such as NeCoRo (Libin and Cohen-Mansfield, 2004, see Figure ??), AIBO (Banks et al., 2008, see Figure ??) or the seal shaped therapeutic robot called PARO (Chang and Šabanović, 2015, see Figure ??). Especially, these cute robots seem to have a great positive influence on the interaction with the elderly by encouraging social interactions, reducing their stress and loneliness and improving their mood (Libin and Cohen-Mansfield, 2004; Šabanović et al., 2013; Aminuddin et al., 2016).

Moreover, SARs are also used as therapeutic tools to support people suffering from cognitive or social disorders, in particular, Autism Spectrum Disorder (ASD) (cf. Dautenhahn and Werry, 2004). Here, they are applied to support the development of life skills to allow for higher independence and to reduce behaviors, which might interfere with this goal (cf. Begum et al., 2016). One of the most prominent aspects patients with ASD are often struggling with is to establish and maintain a social interaction with other people and, hence, this is one of the major topics addressed with SARs. Vanderborght et al. (2012), for instance, used the robot Probo (see Figure ??) to increase the social skills of

[2] https://www.softbankrobotics.com/emea/en/robots/nao

[3] https://www.cmu.edu/cmtoday/issues/dec-2004-issue/feature-stories/human-health/index.html

[4] http://parorobots.com

[5] http://www.megadroid.com/Robots/necoro.htm

[6] https://us.aibo.com/

[redacted]

[redacted]

[redacted]

Figure 3.4: Socially Assistive Robots used to support people with cognitive or social disorders.

children suffering from ASD within a story telling setting and, indeed, their results show that Probo is able to improve their social performance in specific situations (Vanderborght et al., 2012). Another robot that is applied in this scope is called Keepon (see Figure ??). Although it is very simplistic and only capable of expressing its attention, as well as basic emotions, such as pleasure and excitement, it is able to enhance the social abilities of ASD children (Kozima et al., 2007).

Probably the most famous example within ASD research is the robot called KASPAR (see Figure ??). It was developed explicitly for the purpose of supporting children with ASD and according to professionals, it yields a high potential for a broad range of therapies and educational goals in this scope (Huijnen et al., 2016). In fact, different studies have already shown that KASPAR can support children suffering from ASD to develop social and communicative skills, as well as to explore and share their basic emotions (Robins et al., 2012; Wainer et al., 2014). Furthermore, the developed skills were shown to last over a longer period of time, even when the robot is gone again (Robins et al., 2005; Wainer et al., 2014). But, the development of SARs applicable in this field is still ongoing and new concepts are under development, e.g., to relieve therapists or experimenters by increasing the robot's autonomy during the interaction (Zaraki et al., 2018). In addition, concepts for enriching the diversity of robots in ASD therapy can be found, which focus on robots that are already frequently applied in other fields, such as Nao (So et al., 2019; Yang et al., 2019).

3.3 Socially Assistive Robots in Educational Settings for Children

The found positive effects of SARs supports the growing popularity of this field and explains the interest in transferring them to other social interaction settings, e.g., the education of children. But working with a special target group, such as kindergarten children, yields a new set of problems to address, e.g., low automated speech recognition (ASR) accuracy (Kennedy et al., 2017b), a short attention span (cf. David Cornish et al., 2009, p. 73), special interaction dynamics (cf. Lemaignan et al., 2018) and the

[7] http://probo.vub.ac.be/Probo/

[8] http://www.herts.ac.uk/kaspar/the-social-robot

[9] https://beatbots.net/my-keepon

[redacted] [redacted] [redacted]

Figure 3.5: Socially Assistive Robots used in cHRI.

evaluation of systems while not being able to use traditional questionnaires (cf. Belpaeme et al., 2013a). As a result, the novel field of Child-Human-Robot Interaction (cHRI) was established in recent years (Belpaeme et al., 2013a).

One of the major aspects to consider when building a cHRI system is the difference between how children and how adults view the world, especially in their perception of robots. While adults mostly interact with robots quite carefully, children tend to accept them as peers (Tanaka and Matsuzoe, 2012a) or even connect with them on a friend-like basis (Kanda et al., 2004). In fact, even a rather simple and small robot, such as Keepon (see Figure ??), was shown to be able to build social bonds with children just by using simple non-verbal behaviors, such as eye contact, joint attention and basic expressions of affect (Kozima et al., 2009). Especially this aspect gets important when facing the new set of problems and trying to employ a SAR to support children in a variety of educational settings. It cannot only simplify the process of establishing the robot as a social interaction partner, but can also result in a higher enjoyment of the interaction (Shahid et al., 2011) and boost the children's learning gains (cf. Belpaeme et al., 2013b). Consequently, the social bonding might be able to cushion of some of the inconsistencies arising from the new set of problems to address when working with children, which allows to make use of the remaining positive influences of SARs and, with that, to improve educational settings for young children.

A lot of studies have already investigated how SARs can extend ITSs to enrich the tutoring interactions for children and the evaluations of these so-called SARTSs (cf. Clabaugh et al., 2015) highlighted the robots' ability to provide a lot of benefits in a variety of learning settings (e.g., Feil-Seifer and Mataric, 2005; Saldien et al., 2006; Leite et al., 2013; Clabaugh et al., 2015; Kennedy et al., 2015; Gordon et al., 2016). For example, a SAR can be used as an extension to the *user interface* of an ITS to provide engaging long-term guidance and support (cf. Leite et al., 2013), while being able to individualize the one-on-one tutoring interaction for the child learner (cf. Kennedy et al., 2015; Gordon et al., 2016). Furthermore, in this field of research SARs can also take a wide variety of physical forms. They range from pet-like companions, such as Aibo (see Figure ??) or Pleo (see Figure ??), which were success-

[10] https://beatbots.net/my-keepon

[11] https://us.aibo.com/

[12] https://www.pleoworld.com/pleo_rb/eng/index.php

fully used in cHRI, e.g., to support children with health care issues (Saldien et al., 2006) or cognitive disorders (Feil-Seifer and Mataric, 2005), up to humanoid robots, such as Nao.

While the development of SARs is proceeding, they show increasing social-cognitive awareness (cf. Belpaeme et al., 2015) and the applied adaptation methods become more and more sophisticated (see Section 3.3.2). However, since children tend to easily establish a social bond with the robot, it is also important to consider carefully how to introduce the robot.

Different studies already showed that robots can act as a tutor, which supports the learner with appropriate hints (Leyzberg et al., 2012) and lessons (Kennedy et al., 2015). But it can also take the role of a peer, which offers the possibility for the robot to learn together with the child, e.g., in a learning by teaching setting (Tanaka and Matsuzoe, 2012a; Lemaignan et al., 2016b), or as a peer-like tutor that can be regarded as a mixture of both, i.e., as a peer that uses pedagogical well-established strategies to scaffold learning (Belpaeme et al., 2018b).

All three types to introduce the robot were recently used in different projects. In the Emote[13] project, for instance, the robot Nao acted as an emphatic tutor to teach map-reading skills. In the ALIZ-E and CoWriter[14] projects, however, the robot is introduced as a peer, who either supports children in handling their diabetes or needs to be supported with its writing skills in a learning by teaching fashion. In contrast, the L2TOR project introduced the robot as a peer-like tutor and applied it to support young children's language education.

In summary, SARs can take different roles in tutoring interactions with children. However, each role provides its own benefits so that the selection of a role requires careful consideration during the design process of a SARTS.

3.3.1 Language Learning with Socially Assistive Robots

Research already highlighted the capabilities of SARs to provide social and beneficial tutoring interactions in a variety of educational settings, but it often focuses on simple learning domains. For example, when children start to learn mathematics they have to learn simple skills, such as adding, subtracting or multiplying numbers. Language learning, however, provides more complex and, in particular, strongly dependent concepts already from the beginning. For example, learning to apply the grammar of a language requires a sufficient vocabulary. Learning to read and write novel languages requires reading and writing skills in general, which are also not established in kindergarten age. But still, many studies have started to investigate how SARs can be applied to language learning in the past decades.

To tackle the complexity problem, they often break down the large domain of language learning into smaller parts and focus on just one specific aspect. This results in a broad range of addressed subtopics, such as grammar learning (e.g., Herberg et al., 2015; Kennedy et al., 2016), learning how to read (e.g., Hyun et al., 2008; Gordon and Breazeal, 2015; Hsiao et al., 2015) or speak (e.g., Lee et al., 2011; Rosenthal-von der Pütten et al., 2016) and sign language (Uluer et al., 2015), while most studies focus on word learning (e.g., Tanaka and Matsuzoe, 2012b; Alemi et al., 2014; Mazzoni and Benvenuti, 2015;

[13] http://www.emote-project.eu/

[14] https://chili.epfl.ch/cowriter

[redacted] [redacted] [redacted]

Figure 3.6: Socially Assistive Robots used in language education.

Westlund et al., 2015; Gordon et al., 2016; Westlund et al., 2017). In addition, SARs were also used to motivate the learner (Han et al., 2008; Alemi et al., 2015, 2017), to investigate its novelty effect (You et al., 2006; Rintjema et al., 2018) or the effect of its social behaviors (Saerbeck et al., 2010; Haas et al., 2017) in language learning settings (see van den Berghe et al. (2019) for an overview). The frequent application of SARs in this topic can be explained by the benefits a robot provides compared to classical on-screen agents. First, its social presence can help to establish a social interaction, which was argued to be an important prerequisite for language learning (see Section 2.2). Second, a robot is able to interact within real-life environments, which has shown to be another important factor for language development in childhood (cf. Hockema and Smith, 2009). In fact, studies have shown that the manipulation of physical objects (Kersten and Smith, 2002) as well as the application of gestures or even whole body movements (Rowe and Goldin-Meadow, 2009; Mavilidi et al., 2015; Toumpaniari et al., 2015) are able to support children's vocabulary learning.

While the variety of applied robots is as broad as the addressed topics, their provided modalities also vary widely. A famous example is the humanoid robot Nao, which can rely on human-like body parts, such as legs, arms and hands, to express itself in a teaching interaction. Alemi et al. (2014), for instance, employed Nao as a teaching assistant in an English lesson for children. Its presence resulted in a positive impact on their learning gains and speed compared to a control group without a robot assistant (Alemi et al., 2014). Moreover, Kennedy et al. (2016) used a Nao to investigate whether social aspects of a tutoring robot's speech can influence children's second language learning gains, which, however, showed no significant differences for any of their manipulations (Kennedy et al., 2016).

Positive impacts on language learning are also observed for non-humanoid robots, such as Tega (see Figure ??) or Dragonbot (see Figure ??), which resemble fantasy-like creatures. Gordon et al. (2016), for instance, used Tega in a word learning session with personalized affective feedback. They showed that all children learned several words from the interaction with the robot, although no significant differences between the conditions were found (with or without personalization). But more importantly, children felt more positive towards the personalized robot (Gordon et al., 2016), indicating the

[15] http://robotic.media.mit.edu/portfolio/tega/
[16] http://robotic.media.mit.edu/portfolio/dragonbot/
[17] http://inrobotek.com.tr/ProductWithTab.aspx?MenuID=27

robot's effects onto the affective learning dimension, which in turn can lead to a higher long-term motivation and, thus, to repeated sessions of language learning assisted by the robot. This finding is also supported by Westlund et al. (2015), who used Dragonbot for word learning and compared a robot tutoring session with a session provided by a human or a tablet. Although their results showed no significant difference in word learning, the children preferred learning with the robot (Westlund et al., 2015). Thus, besides the potential positive effect on children's long-term motivation, their results further provide an indicator that a robot can be as effective in teaching simple vocabulary to children as a human teacher. A third example is the robot iRobiQ (see Figure **??**), which appears more futuristic compared to the previously presented robots and includes a small touch screen on which the learning content is presented. It was applied in reading exercises during storytelling and demonstrated a positive impact on reading, understanding, re-telling and creation of stories, as well as word recognition abilities compared to a traditional media-assisted reading program (Hyun et al., 2008; Hsiao et al., 2015).

However, only a small group of studies addressed the target group of young kindergarten children (4-6 years) and often investigated more general aspects and effects of SARs. In particular, they examined the effect of their social presence (Westlund et al., 2015), studied whether a child can learn as much from a robot peer as from a child peer (Mazzoni and Benvenuti, 2015) or whether a learning by teaching setting with a robot as a less knowledgeable peer is feasible and beneficial for children (Tanaka and Matsuzoe, 2012b). Finally, it has also been examined whether a robot can generally act as a tutor driven by an adaptive system (Gordon and Breazeal, 2015) or whether the SAR's effects on learning, motivation and engagement found so far are just based on a novelty effect (Rintjema et al., 2018). Just a small portion investigated a SAR's behavioral possibilities in more detail. The corresponding studies included strategies, such as motivational prompts executed by a SAR, to lower anxiety and rise motivation (Gordon et al., 2016; Alemi et al., 2017) or they examined the understandability of human-like nonverbal behavior transferred to a SAR (e.g., eye gaze or body orientation towards an unfamiliar object) (Westlund et al., 2017). But still, the body of detailed knowledge about which actions a SAR can apply to address specific aspects of a learning interaction, let alone a specific dimension of learning, is very limited, especially for young kindergarten children.

3.3.2 Adaptation in Socially Assistive Robot Tutoring Systems

The presented results above show that SARs can have a great potential to increase the learning gains of children and their motivation to continue learning in a variety of topics. To use their full potential the interaction needs to be adapted and personalized to the needs and preferences of each individual child. Especially since students search for less help from a robot compared to a human (Serholt et al., 2014), although both types of learning interactions can still be successful (Huskens et al., 2013), a SAR has to react pro-actively throughout the interaction to provide the required support for each child. In order to allow for the tutoring system to adapt appropriately, further information about the learner's cognitive and affective state is required, and, thus, has to be tracked by the system. But this is a complex task, especially since children, as humans in general, possess a lot of individual differences, which have to be considered, as well. They include not only general aspects, such as different levels of social skills, uptake abilities or attention spans (cf. Hooper and Umansky, 2009, Chap. 1; David Cornish et al., 2009, p. 73), but also elements specific for language learning. For example, the development of different neuronal patterns is dependent on the age in which either a second or a foreign language is learned (see Section 2.2.1). These patterns can influence the language learning later on and, thus, result in individual differences between children. In addition, longitudinal studies showed that differences in early mother tongue skills (e.g., phonological awareness and word decoding) can also influence the learning of a second or foreign language later on. However, this applies only for learned language skills (Sparks, 2012), while skills that are based on genetic factors and contribute to second or foreign language learning seem not to be affected (e.g., listening and responding, speaking or writing) (Dale et al., 2012). Nevertheless, the language skills learned in the first years of life can shape children's abilities so that they learn slower or faster as usual or even develop a deficit in specific language skills. Consequently, to enable a SARTS to adapt the tutoring interaction appropriately, no standardized solution, which assumes that every child learns the same, can be used and sophisticated tracking algorithms are required. With these, the tutoring system is able to tailor the interaction to the learner's individual needs, which in turn will result in a better learning progress (Leyzberg et al., 2014; Gordon and Breazeal, 2015). Consequently, the multidimensionality of learning yields a lot of possibilities and necessities for adaptation, but increases the complexity enormously.

Since the goal of each tutoring interaction is to teach new knowledge or skills, the first and most obvious adaptation possibility addresses the cognitive learning dimension and is often based on the learner's task performance during the interaction. For example, a SARTS can provide personalized hints for the learner, which has been shown to result in a more successful interaction, a reduction of time needed for problem solving and a higher motivation (Leyzberg et al., 2014). Another option is the personalization of task feedback based on children's task performance, which has been shown to result in a more effective and less frustrating interaction with a Nao robot (Greczek et al., 2014). Other research focuses on adapting the lesson (e.g., Leyzberg et al., 2018) or task order (e.g., Käser et al., 2014a) to precisely address the learner's weaknesses, or the learning style to match her preferences (e.g., Clabaugh et al., 2015). In summary, a lot of promising evidence can be found that adaptation in the cognitive learning dimension is feasible and can be highly supportive and beneficial for the learner.

Another important aspect besides addressing the cognitive learning is the adaptation based on the learner's affective learning or rather her engagement. Szafir and Mutlu (2012), for instance, adapted the robot's gestures and speech based on the engagement level measured with an EEG. Their results show that the adaptive robot behavior during a storytelling session raised the recall abilities of the learner afterwards (Szafir and Mutlu, 2012). Although techniques like EEG or similar tracking possibilities are often quite precise in keeping track of the learner's affective and cognitive states, they are also very intrusive, e.g., by requiring a lot of wires. However, an increasing body of research also investigates how interactions can be personalized to the affective states based on less intrusive tracing technologies (Jones et al., 2015; Leite, 2015; Ramachandran and Scassellati, 2015). For example, Gordon et al. (2016) demonstrated that adapting the interaction based on an off-the-shelf affect tracking framework called Affectiva Affdex (McDuff et al., 2016), which works with a common webcam, can already support the learner. Although, they used just a simple personalization strategy based on reinforcement learning, starting from simply mirroring the learner's affective states, it already resulted in an overall higher valance and, thus, positive affective states (Gordon et al., 2016).

The last dimension to address is the perceived learning of the student. But this dimension is often not addressed with the intention of evoking positive influences on the other dimensions, but to support a process called SRL (cf. Schunk, 1987; Schunk and Zimmerman, 2007). Within this concept, the learner adapts her learning interaction herself based on an estimation of her own knowledge. Thus, to allow for a SARTS to support this process it not only needs to keep track of the skill mastery but also requires the ability to communicate this knowledge. Opening up this internal knowledge base is commonly referred to as Open Learner Model (Bull and Kay, 2010) or system transparency (Lyons, 2013; Mercado et al., 2016; Lyons et al., 2017), which has been shown to help students to better regulate their efforts (Bull et al., 2010) or to improve their problem selection (Mitrovic, 2010). Additionally, transparency about the system's states improve trust in the system as a whole by making its behavior more understandable (Lyons, 2013; Mercado et al., 2016; Lyons et al., 2017) and, therefore, reducing uncertainty in the user. In general, Open Learner Models can take the form of a series of skill meters (Bull et al., 2010; Long and Aleven, 2013; Jones et al., 2017; Jones and Castellano, 2018), but the knowledge can also be expressed verbally, e.g., by a robot (Jones et al., 2014). Jones and Castellano (2018), for instance, visualized the estimated skill knowledge of the learner during map-reading tasks via a skill-meter on a screen and used a Nao robot to adaptively scaffold the learner's SRL process. Their results show that a SAR is not only able to support the process in general but also to improve the learner's individual abilities of SRL compared to the control condition without robot support. However, they were not able to find significant differences between the conditions, although the learners had a slightly higher learning gain in the robot supported condition (Jones and Castellano, 2018).

3.4 Summary

To build and structure a SARTS the concepts and implementation paradigms of ITSs can serve as a basis. The goal of ITSs is to apply artificial intelligence to educational settings and to provide a set of standard components. Although the commonly used architecture presented by Dede (1986) consists of four modules, definitions including only three can be found as well. However, all architectures have the same set of modules in common: a *domain model* to provide the curriculum, a *student model* to manage information about the learner and a *pedagogical module* that plans the next steps of the tutoring interaction based on information of the *domain* and *student model*. The extended architecture of Dede (1986) further includes a *user interface*, which is used to communicate with the learner.

In addition to the basic architectures, three major implementation paradigms can be found in the literature. While Rule-Based Models (RBMs) focus on modeling the whole solution process step by step, Constraint-Based Models (CBMs) use simple rules that just have to be satisfied by the final solution. Although this simplifies the modeling task, it also results in a loss of detail about the learner's proficiency. Finally, the third approach is based on incorporating an expert system into an ITS, which can be used in two different ways. First, a variety of solution paths can be generated based on the expert system and, subsequently, consulted to provide useful hints or to compare them with the learner's answer. Second, the expert system can be used to generate just the ideal solution, which is compared to the learner's answer, and to provide detailed feedback about possible problems. Consequently, choosing the right architecture and paradigm to implement an ITS is strongly task and domain dependent and, thus, needs careful consideration during the development process.

However, when designing an ITS for language learning a social component is recommended, which allows for social and personalized interaction by offering motivational and engaging long-term support. Here, the physical and social presence of SARs, which already demonstrated their benefits in a variety of settings, can be used to create a SARTS, which further allows to establish a so-called *pragmatic frame*. Although this concept can also be used in multiparty interactions, a SAR also provides the option to establish personalized one-on-one tutoring interactions, which offer the potential to overcome the weaknesses of usual classroom instructions. To achieve this, a SARTS has to adapt the interaction with respect to the learner's individual needs in each learning dimension, for example, by detecting and interpreting children's behavioral cues that can be used to identify problems or negative affective states during learning, such as boredom or frustration. Further possibilities are to track learners' knowledge state and preferred learning style. Subsequently, all this information can be used to adapt the interaction to the individual needs and preferences of each learner to provide an optimal tutoring experience.

In conclusion, the reviewed literature provides general concepts regarding the required modules to develop an ITS or SARTS, respectively, as well as general paradigms to follow learners' solving process and to validate their answers. In addition, information on how a SAR can be used to interact with humans in a variety of settings was collected, which can inform the development of novel scaffolding strategies for a SARTS. But for selecting a suitable ITS paradigm and architecture, as well as a role for the SAR, further information on the general setting and structure of the tutoring interaction, as well as the applied feedback behavior is required. Further, these aspects have to be suitable for young kindergarten children to establish an optimal learning environment for teaching a foreign language to them and, therefore, they are addressed in the following chapter.

Children learn as they play. Most importantly, in play children learn how to learn.

— O. Fred Donaldson

4 Designing the Tutoring Interaction

After the general paradigms and architectures of ITSs, how they can be enriched by SARs and which benefits they provide were summarized, this chapter focuses on finding a suitable design for a language tutoring interaction that can be provided by a SARTS. It should allow for a SARTS to provide multidimensional support for kindergarten children's language learning. Further, it has to provide the necessary information to answer the open questions of which role the SAR should take and which ITS paradigm and architecture should be chosen.

To inform this, observational recordings of language learning interactions in German kindergartens were collected and analyzed (Section 4.1). Although these interactions mainly took place in a group-like fashion, the recorded data still provide valuable hints regarding an appropriate scenario and structure that can also be applied in one-on-one tutoring interaction driven by a SARTS and, with that, allows to further study the effects of SARs for foreign language learning (**RQ0**, Section 4.2). Based on the derived interaction design, a general tutoring system is designed (Section 4.3) and implemented into a modular technical system (Section 4.4), which serves as a basis for all evaluation studies conducted within the scope of this thesis.

4.1 Empirical Basis

To design and implement a tutoring interaction suitable for a SARTS that matches kindergarten children's needs, an empirical basis of language learning interaction data is required. The goal is to examine whether language learning practices in kindergartens contain elements that can be transferred and implemented into a SARTS (**RQ0**). To this end, video recordings of language tutoring games, as they take place in German kindergartens, were collected. Since one-on-one interactions between an educator and a child are hardly realizable in kindergartens, the games typically involve one educator and a small group of children.

4.1.1 Video Data

In total, the collected dataset comprises about 681 minutes of video data with an average video duration of 22:20 minutes ($SD = 1{:}17$ minutes). The recorded interactions contain information about four different learning games including three card games called “I spy with my little eye ...” (Figure 4.1a), “I am giving you a present ...” (Figure 4.1b) and a card-based rhyming game (Figure 4.1c), as well as the reading of a picture book in an interactive manner (Figure 4.1d). The two observed rhyming game lessons took place in one-on-one interactions between one child and an educator, whereas in the other three games the educator was playing with three different children. The recorded children were between the age of 4 and 6 years and learned German as a second language, while already knowing basic vocabulary. Furthermore, the educators are of different age with a varying amount of working experience, which allows to get a broader overview about possible teaching practices.

4.1.2 Analysis

To be able to derive useful information for designing a SARTS the recorded dataset is transcribed and annotated with regard to the following categories:

- **Dialog acts**: Utterances are classified with respect to the underlying intention based on the Dialog Act Markup in Several Layers (DAMSL) annotation scheme (Core and Allen, 1997).
- **Children’s mistakes**: Types of language errors the children made, e.g., wrong plural form, missing articles, wrong syntax, etc.
- **Educator’s repair strategies (feedback)**: Pedagogical acts used to correct the errors, e.g., reformulation, corrected repetition, etc.
- **Nonverbal behavior**: Nods, smiles and gestures used by the educators.

4.1.3 Observational Results

The analysis of the annotated dataset revealed general patterns that can be used to design a SARTS. Basically, these patterns can be divided into three categories: (1) a general interaction structure, (2) the educator’s feedback behavior and (3) settings for the tutoring interaction.

4.1.3.1 General Interaction Structure

Analyzing the recorded learning session revealed a common structure used in all interactions.

1. **Opening:** This phase marks the beginning of the interaction and is intended to motivate the child and to mitigate her timidity. Usually this is done by inviting the child to introduce herself or by joint singing of a welcome song.
2. **Game setup:** This step is used to prepare the game by explaining the task and clarify the necessary terms. It is usually done in the language to be learned, here German, but only if some basic vocabulary is already known.

3. **Test run:** A test task is presented to practice the game flow and to check whether the game instructions have been understood. This is further used to reduce children's pressure and, again, to mitigate their timidity.
4. **Game:** Here, the main part of the learning interaction takes place. Depending on the game, an object is explained or a question is asked based on the presented material, while every move or answer of the child is accompanied by the educator's feedback and motivational prompts.
5. **Closing:** This phase marks the end of the learning interaction. Additionally, it is used to maintain children's motivation for future interactions by acknowledging their participation, joint singing a goodbye song and providing an outlook on what is going to happen next time.

4.1.3.2 Educator's Feedback Behavior

In addition to the general interaction structure, also the educator's behavior when providing feedback to the children was analyzed. An important and commonly used pattern is that language errors are almost never marked as wrong or explicitly corrected. Instead, feedback is always provided in a positive way falling into one of the following four categories with the percentage of their occurrence given in squared brackets:

1. **Implicitly correcting** the child after a mistake, i.e., repeating the correct word as if it was already used correctly (e.g., correct pronunciation, with article, plural form, etc.) [54%]
2. **Correctly recasting a sentence**, e.g., after syntax errors [32%]
3. **Praising the child** for a correct answer, which is often combined with a repetition of the correct utterance [13%]
4. **Moving on to the next task without corrections**, e.g., when children's message is unclear due to incomprehensible pronunciation [1%]

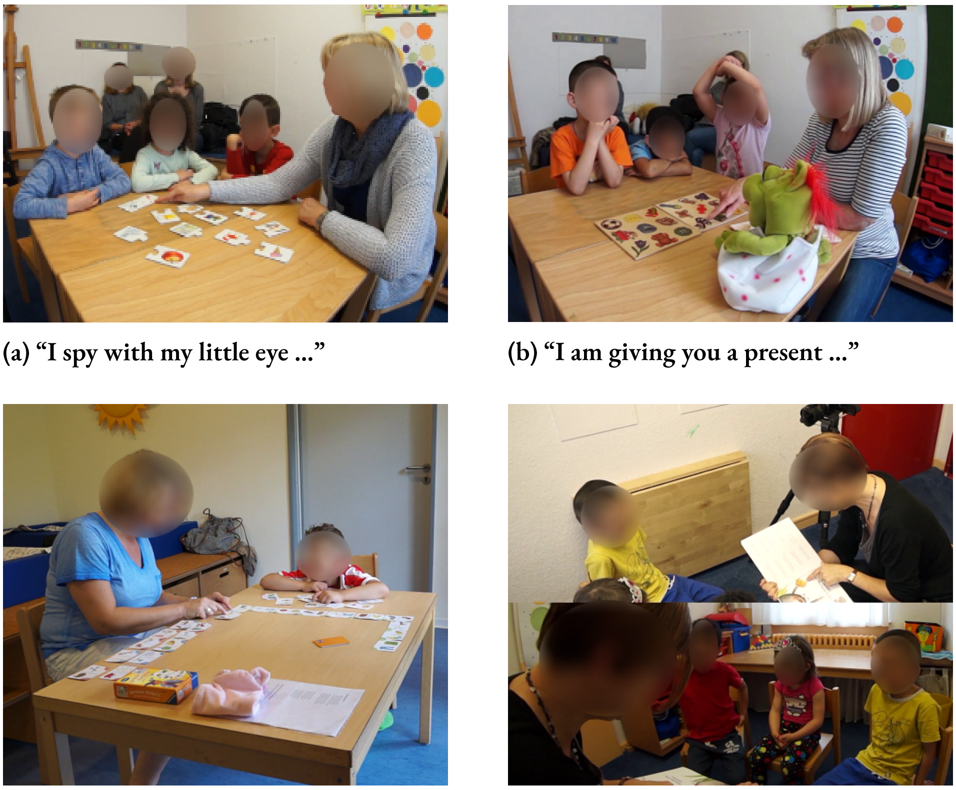
Furthermore, educators' feedback behavior is typically accompanied by socially supportive nonverbal behaviors, such as smiling and nodding.

4.1.3.3 Language Learning Games

Since all observed tutoring games follow the same interaction structure described above, they mostly differ in their the game phases.

In the "I spy with my little eye ..." game a player describes an object presented on a card that lies between different distractors on a desk (see Figure 4.1a). To describe the object, the player has to use adjectives or subjects from the to be learned language. Subsequently, the other players have to guess or preferably know which object is described and should point at it.

Although the "I'm giving you a present ..." game (see Figure 4.1b) is similar to the "I spy with my little eye ..." game in its description phase, its answer mechanic is different. One player picks a card with an object not yet visible to the other players and starts to describe it. But this time, the other players have to guess which object it is without knowing which objects might exist within the game context.



(a) "I spy with my little eye ..."

(b) "I am giving you a present ..."

(c) Rhyming game

(d) Reading a picture book

Figure 4.1: Image cutouts from four different language learning games in German kindergartens.

In the rhyming game, each player has the same number of cards laying on a table in front of them (see Figure 4.1c). At the beginning, one card that displays two different objects is laying in the middle of the desk. Now, the players have to find a fitting card that displays an object, whose name rhymes with one of the two objects on the card in the middle. If they find a suitable card, they have to say the object names before putting it next to the card in the middle, with the rhyming objects side by side.

In the last game, the reading of a picture book, the educator asks the children to explain what they see on the different book pages, while she is mainly listening (see Figure 4.1d). Only if the children do not know how to go on or a child makes a mistake, the educator is intervening and either helps with some hints or corrects the child's utterance.

4.2 General Interaction Design

The information of the empirical basis helps to develop a suitable tutoring interaction for kindergarten children that can be provided by a SARTS to support foreign language learning. To this end, the following aspects are considered carefully. First of all, a role for the robot needs to be selected with respect to the optimal set of benefits for language learning. Furthermore, the game setting, structure and feedback behavior for the tutoring game have to be chosen and adapted to a cHRI, which are informed by the useful insights summarized in the empirical basis.

4.2.1 The Robot's Role

As described in Section 3.3, three different roles for the robot are commonly used in cHRIs. First, the robot can take the role of a tutor, whose main task is to support the learner during the interaction, e.g., with appropriate hints and personalized lessons. However, this might cause higher expectations in the robot's abilities and competences. Introducing it as a peer, instead, could raise the acceptance of minor problems, such as a suboptimal interaction flow due to technical issues or limitations of the robot, e.g., slow reactions because of difficulties to interpret children's behavior, and can help to maintain children's social bonding (cf. Belpaeme et al., 2018b; Lemaignan et al., 2015). Finally, the robot can also be introduced as a peer-like tutor that can be regarded as a mixture of both, i.e., as a peer that uses pedagogical well-established strategies to scaffold learning (cf. Belpaeme et al., 2018b).

With respect to the subject of language learning, the role of a peer-like tutor fits the requirements best. First of all, it is also the commonly used role taken by the educators in the observed tutoring games in German kindergartens and, thus, can be assumed to be suitable for each game. Second, this role does not only enable the robot to guide the child through the interaction as a tutor but also to introduce itself as a peer who wants to play a game together with the child. This allows to also benefit from the positive effects of the peer-role, which can support the establishment of a solid social bonding and, with that, can help to maintain a high long-term motivation. Furthermore, children might forgive smaller technical problems, which can occur during the interaction, e.g., due to limitations of the robot.

4.2.2 Setting for the Tutoring Game

As discussed in Section 3.3.1, learning a language is a complex task with strongly dependent concepts right from the beginning. This is the reason why often just a small part of this process is considered for the development of tutoring systems. Most of them focus on word learning, which aims for establishing a solid vocabulary. In general, word learning just addresses the *factual knowledge* and the cognitive processes of *remembering*, *understanding* and *applying* so that it should be manageable for each kindergarten child (cognitive learning, see Section 2.1.1). Furthermore, since it can serve as basis for advanced language learning later on, it is a reasonable starting point for teaching kindergarten children a foreign language and this thesis sticks to it as well. Of course, a few children probably already know first words of other languages, but a careful selection of the target vocabulary still enables the children to learn new words.

To be in line with language education applied in German kindergartens, one of the observed games in the recorded dataset can be adopted, since they all provide the potential to create a basic word learning interaction. However, not all of them fulfill the requirements for the implementation into a SARTS. First of all, the tutoring game needs to offer the possibility to adapt the learning content, i.e., word order and task difficulty, as required for tailoring the interaction to the individual knowledge of each learner and, with that, to optimally address her cognitive learning. While all of them offer the option to modify the order of words to be learned, half of the games lack the ability to provide different task difficulties. While this is easily possible during the reading of a picture book or in the

"I spy with my little eye …" game, e.g., by varying the number of distractors on a book page or a table, the remaining two games do not provide this option and, thus, just allow for a SARTS to create a less personalized tutoring experience.

The second aspect to consider is the transferability of the tutoring game to a cHRI driven by a SARTS. Since the reading of a picture book mainly focuses on children's abilities to build complete sentences, although it can be adapted to single word answers, it still requires verbal input. This also applies for the "I'm giving you a present …" game, which requires the speech input to understand either the child's description of an object or her answer to the educator's/SARTS's description. However, this is not realizable with kindergarten children, since state of the art ASR systems still yield low accuracy for them (Kennedy et al., 2017b). In contrast, the rhyming and "I spy with my little eye …" games allow for alternative input modalities, e.g., via touch on a tablet, and, thus, can be transferred to a cHRI in a reasonable fashion.

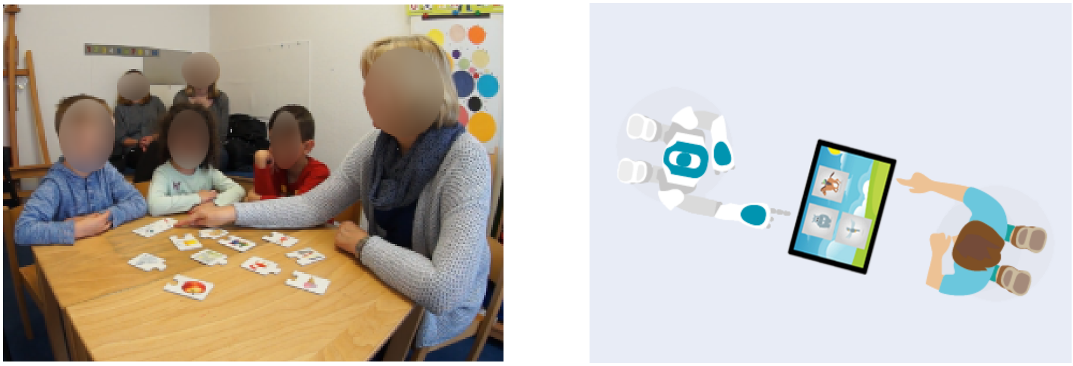
In summary, the "I spy with my little eye …" game meets all requirements and, thus, provides a promising basis to implement a word learning interaction for kindergarten children. It not only yields a good transferability but also allows to easily adapt the interaction with respect to the cognitive learning dimension by modifying the order of words to be learned, as well as the task difficulty. Combined with an approach to keep track of the learner's knowledge state, this offers the possibility for the SARTS to provide an interaction within the ZPD for each individual child. However, to allow for the chosen game setting to be a meaningful aid to learn new languages, some small adaptations are required.

First, the game has to be transferred from an interaction between the educator and multiple children to a one-on-one interaction between the SAR and just one child. Therefore, the SAR takes the educator's role of a peer-like tutor that assists the child in learning novel vocabulary of a new language. However, this also involves that the robot has to handle and understand a lot of speech input from the child, since the educator also took part in the actual game. In order to minimize the required verbal input, the usual role switch during the game is avoided and the robot continuously acts as "the spy", who is describing the game objects by using the target vocabulary.

The second aspect to be changed concerns the visualization of the game itself. Instead of physical cards on a table, a tablet pc is used to display the objects associated with the current task, which further simplifies the implementation and maintenance of the tutoring interaction (see Figure 4.2). Although research indicated that the manipulation of real objects is important for language learning (Kersten and Smith, 2002), it has also been shown that the interaction with objects on a tablet can work similarly well (Vlaar et al., 2017).

4.2.3 Interaction Structure and Feedback Behavior

Despite these changes, the game structure is not changed and still follows the previously identified interaction structure making up a typical tutoring game in adult-child interaction (see Section 4.1.3.1). First, the robot introduces itself and asks the child for her name and age, which is intended to mitigate her timidity for interacting with the robot (opening phase). Afterwards, it explains the game rules



(a) (b)

Figure 4.2: (a) A scene from the "I spy with my little eye ..." game from the pre-recordings in German kindergartens and (b) the setting transferred to a SARTS with a child sitting in front of a tablet displaying the graphical user interface (taken and redesigned with permission from Schodde et al. (2019)).

(game setup) and provides a test task to check whether the child has understood the game (test run). Subsequently, the actual game starts, in which a basic turn is structured as follows.

It starts with a set of objects being displayed on the tablet screen and the robot saying "I spy with my little eye something that is ..." followed by the target word in the foreign language that refers to a property of one of the objects displayed on the screen. Then, the child's task is to respond by selecting the object she thinks is referred to via touch input on the tablet. After the child answered, the robot provides feedback in response to a correct or false answer. Since appropriate feedback is not only important for the child to understand her mistakes but also to motivate her and foster her perceived learning (see Section 2.1.5), the robot's feedback behavior is designed in line with experts' practices summarized in the empirical basis (see Section 4.1.3.2). However, the two most commonly used feedback strategies are not applicable in this context, since the current interaction design restricts verbal input. Additionally, the last strategy of providing no correction for the child is rarely used in general and never in the chosen game. Consequently, the third strategy is applied by the robot so that it responds to correct answers by praising the learner, as well as repeating the target word in the foreign language and the corresponding translation in the mother tongue. In case of a false guess the robot explains the correct meaning of the word to be learned one more time. In addition, the wrongly chosen and the correct object are both displayed on the tablet screen and the child is again requested to provide an answer. Similar to the educators, the robot also accompanies all feedback behaviors with small socially supportive behaviors, e.g., nodding when an answer is given. Finally, the interaction is closed by acknowledging the child's participation and providing an outlook for the next session.

In general, combining the identified interaction structure with an established way to provide feedback for young children, guided and performed by a SAR, facilitates the establishment of a *pragmatic frame*, as well as a social interaction between the learner and the robot. This should foster that children get used to the interaction quickly, so that they can fully concentrate on the learning content, while benefiting from the social interaction, which has been argued to be the best setting for language learning (see Section 2.2).

4.3 General Tutoring System

In general, the "I spy with my little eye ..." game implies a communication between the child and the SARTS. To make this possible, the ITS architecture proposed by Dede (1986) is used, since it already includes a *user-interface*. Here, it is represented by the SAR and a Graphical User Interface (GUI) that displays the cards with objects "the spy" can describe during the game. As SAR the Nao robot is used, since its humanoid appearance allows to apply useful behaviors derived from educators' practice in kindergartens. Further, it looks kind and has a suitable size, since it is smaller than kindergarten children so that its movements might not frighten them. To further lower the children's initial anxiety before requesting them to play alone with the robot, it is introduced in a group session (cf. Vogt et al., 2017; Fridin, 2014a).

In addition to the *user interface*, Dede's architecture includes a *domain model*, which is used to specify the to be learned vocabulary, as well as a *student model* and a *pedagogical module*. Within this thesis, the *student model* is used to trace the learner's knowledge state about the vocabulary and her engagement, whereas the *pedagogical module* is applied to plan the next steps of the tutoring interaction based on the *domain* and *student model*.

Since the learning dimensions and the engagement are interconnected and strongly influence each other, this needs to be considered during the planning process of SARTS. Therefore, the tutoring system is developed and implemented in the same fashion by combining the *student model* and *pedagogical module* into a single approach. This does not only allow to consider all the information about the learner during the decision-making process but also to simulate the effects of possible actions beforehand and, with that, to choose appropriate teaching actions in each situation.

Finally, since the "I spy with my little eye ..." game provides such a simple task structure, it further allows to rely on the paradigm of a Constraint-Based Model (CBM) to implement the SARTS (see Section 3.1.2). This is because the tasks require just one-step answers by selecting one of the shown cards and do not provide the necessity to track the learner's solution path for which Rule-Based Models (RBMs) would be needed. Moreover, incorporating an expert system for such a simple task is not required and would result in additional modeling effort.

4.4 Technical Setup

The SARTS consists of a Microsoft Surface Pro 4[1] tablet to display the HTML-based GUI and a Nao robot V5[2] to guide the learner through the interaction. Furthermore, it contains a dialog manager to specify the interaction structure easily and a control panel, which allows to start, pause and stop the SARTS. Finally, the developed SARTS includes a backend that, inter alia, contains the *student model*, as well as the *pedagogical module*.

To allow for an easy exchange of single components with respect to the respective study settings and latest findings, the SARTS is constructed modularly (see red components in Figure 4.3). Furthermore,

[1] https://www.microsoft.com/surface/en-gb/devices/surface-pro-4

[2] https://www.softbankrobotics.com/emea/en/robots/nao

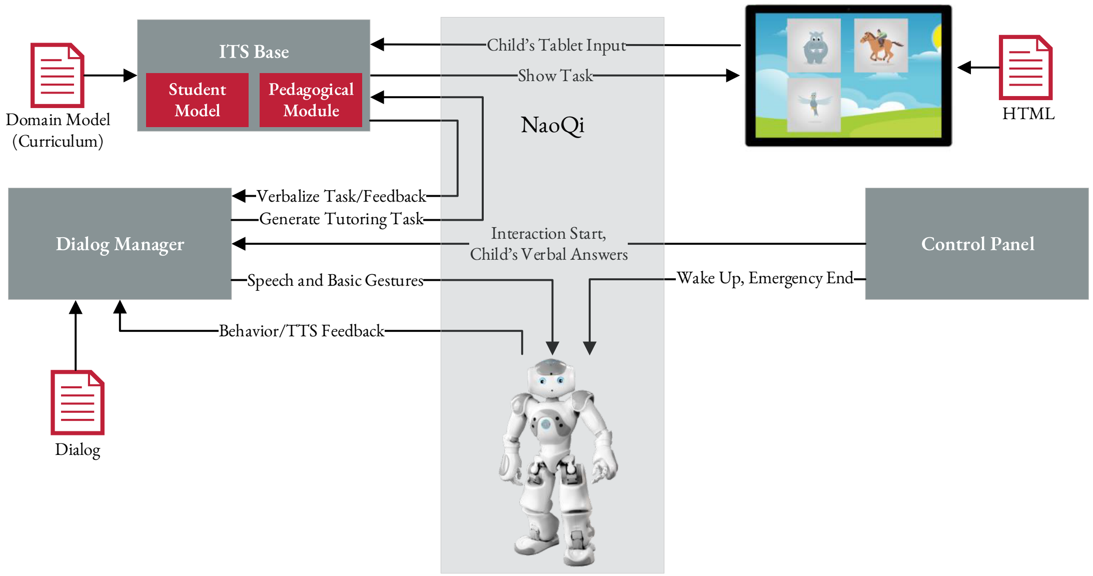


Figure 4.3: Overview of the technical realization of the basic tutoring system. The arrows show the information flow via NaoQi that connects the different modules. The red components can be exchanged or modified for each experiment.

the NaoQi[3] framework is applied as a middleware to allow for all components to communicate with each other. It is shipped with each Nao robot and allows to communicate via a simple event based system between various programming-runtimes (Python, Java, C++, JScript) and, with that, provides a high flexibility.

Figure 4.3 further depicts the information flow between the different components. For example, during a normal task, the dialog manager asks the backend to generate a new task or to select a target word and task difficulty, respectively. Subsequently, the backend informs the GUI about which objects to display and also fills up template slots in the dialog manager, e.g., with the target word. The dialog manager sends this information to the robot to verbalize the task and/or to start non-verbal behaviors. Finally, when the child selected an animal on the tablet via touch, this answer is sent to the backend where it is validated. Afterwards the feedback information is sent back to the dialog manager to fill up the respective dialog templates and the completed sentences are then expressed by the robot.

4.5 Summary

To be able to investigate the important research questions arising in the field of language learning for young kindergarten children, a suitable interaction design is required. To get an impression of which elements of language learning practices in kindergartens exist and which can be transferred and implemented into a SARTS (**RQ0**), observational recordings of language learning interactions in German kindergartens were collected and analyzed. Based on the resulting dataset, an interaction structure, feedback guidelines and a tutoring game are derived and implemented.

[3] http://doc.aldebaran.com/2-1/naoqi/

In general, each of the observed language learning games in German kindergartens follow the same interaction structure consisting of five different phases (opening, game setup, test run, game, closing). Further it has been observed that the educators almost only use positive feedback during the actual game, such as simply correcting children's answers implicitly either by repeating the correct word as if it was already used correctly by the child or even recasting a whole sentence in a corrected fashion. On the one hand, this can keep the children motivated to continue playing the game and, thus, to learn more words. On the other hand, the implicit corrections through repeating, recasting or even rephrasing children's answers correctly results in a higher contact frequency with the target word and, with that, can lead to a higher cognitive learning. Although young children are able to pick up a rough word concept through *fast mapping*, frequent repetitions are still essential to strengthen this initial concept and to enrich it with more details in the *slow mapping* process (see Section 2.2.2).

All these important aspects can be established in the "I spy with my little eye ..." game, which is also easily transferable to a SARTS without the need of excessive turn taking during a task and speech input by the child. Furthermore, it provides the opportunity for the SAR to establish a *pragmatic frame* by acting as the game master in the role of a peer-like tutor that guides the children through the different interaction phases. While this frame makes it easier for children to retrace the interaction course and to concentrate on the learning task itself, it further supports to establish the robot as a social partner and maybe also as a friend.

Afterwards, all the aspects described above have been transferred into a technical setup for building a modular SARTS, which allows to address the remaining research questions of this thesis. Now, some modules of the underlying ITS, such as the *student model* or *pedagogical module* need to be developed and implemented, which will be addressed in the following chapter.

If a child can't learn the way we teach, maybe we should teach the way they learn.

— Ignacio Estrada

5 Scaffolding Cognitive Learning

After a suitable tutoring setting and structure was defined and transferred into a SARTS, this chapter focuses on developing the core modules that are generally required for a SARTS to optimally support the foreign language learning of kindergarten children. More precisely, it focuses on the question of how a SARTS can optimally address young children's cognitive learning (**RQ1**). This dimension comprises the knowledge and skills to be learned and, thus, is directly connected with the interaction goal of teaching new content. To optimally address the cognitive learning and to achieve high learning gains, a SARTS should allow for the learner to work in the ZPD. To enable the system to adapt the learning content accordingly, this chapter concentrates on how a SARTS can be enabled to keep track of kindergarten children's individual knowledge state (**RQ1.1**, *student model*) and how it can select appropriate teaching actions to adapt the interaction accordingly (**RQ1.2**, *pedagogical module*).

To approach these questions, this chapter first reviews the literature on possible models to trace the learner's knowledge, as well as to plan and adapt the tutoring interaction (Section 5.1). Subsequently, and with respect to the reviewed literature, a model is selected (Section 5.1.3), extended, formally defined and implemented (Section 5.2). Afterwards, two user studies are described, which evaluate this model in an interaction with adults (Section 5.3) and with children (Section 5.4).

5.1 Model Selection

A lot of different approaches for tracing the learner's knowledge and, based on that, for planning and adapting the course of the tutoring interaction have been published in the past decades. Since they define the basis for the *student model* and *pedagogical module* of an ITS, they are reviewed and discussed in the following. As a first step to develop a *student model*, this chapter will only focus on modeling the learner's knowledge state, although more information is required to establish a *student model* that profiles the learner completely, e.g., affective and cognitive states, and preferred learning style.

5.1.1 Modeling the Learner's Knowledge State

Approaches of knowledge tracing often aim to model the learner's mastery of skills being taught during a tutoring interaction (see Pelánek (2017) for an overview). This is one of the important pieces of information that is stored in the *student model* of an ITS and can be used as a knowledge base to address the learner's individual needs by planning the next steps in a tutoring interaction accordingly. One possible way to approach this aspect is to extract information about the student from data using complex machine learning algorithms, such as recurrent neural networks (Piech et al., 2015; Khajah et al., 2016), collaborative filtering techniques (Töscher and Jahrer, 2010) or even ensembles of different approaches (Pardos et al., 2012). While these models often achieve a good predictive accuracy, they lack interpretability. However, especially in the scope of educational applications, educators and teachers are often concerned with the interpretability and validity of applied models and, thus, these approaches are barely used in practical applications. Alternatively, simple and easily understandable assumption-free approaches, such as the exponential moving average, can be used. Here, past attempts to solve a task are weighted by an exponentially decreasing function to estimate the learner's knowledge. These approaches have the advantage of computational efficiency and the ease of application, while often providing reasonable predictions. Nevertheless, they still cannot keep up with more sophisticated knowledge tracing algorithms (cf. Wauters et al., 2012; Pelánek, 2014).

A more elaborated approach to model the learner's knowledge is based on logistic models, which are usually used to model the acquisition and forgetting of declarative knowledge (White, 2001; Pavlik and Anderson, 2005; Pelánek, 2015; Sense et al., 2016). To achieve this, the skill is represented as a continuous variable and learning is modeled as a gradual change. Furthermore, the item difficulty is calculated by using a logistic function, e.g., $f(x) = 1/(1 + e^{-x})$, representing the probability of answering correctly given a specific task difficulty and the current skill mastery. A typical logistic model is the Performance Factor Analysis (Pavlik et al., 2009), which allows to estimate the skill mastery based on the learner's performance during the interaction. Similar models are the Additive Factors Model (Cen et al., 2006; Käser et al., 2014b), which is also sensitive to the frequency of prior practices of a skill, the Instrumental Factors Analysis (Chi et al., 2011), which also incorporates different types of instructional interventions and their effects, and the Elo Rating System (Pelánek, 2016). The latter is originally developed to rate chess players and allows to easily and dynamically estimate the skill level of students, as well as the difficulty of tasks by interpreting the student's answer as a match between the student and the task.

Another widely used group of approaches incorporates Bayesian models. They are able to handle uncertainty easily, recover from errors during an interaction and allow to infer hidden state values from evidence. The On-Line Assessment of Expertise (OLAE) tool, for instance, observes the individual steps done by the learner to infer her skill mastery and domain knowledge (Vanlehn and Martin, 1998). Similarly, the ITS called Ecolab logs the learner's requests for help to predict the mastery of the current domain, as well as the readiness to learn new topics (Luckin and du Boulay, 1999). Moreover, Gordon and Breazeal (2015) presented a so-called "active learner model" to trace the word reading skill of young children. It employs a simple distance metric to approximate the conditional probability $p(w_2|w_1)$ that describes whether the child might be able to read a word w_2 if it already knows the word w_1. The evaluation showed that this approach is able to adapt to users of different age and to trace their reading knowledge fairly well (Gordon and Breazeal, 2015). An additional benefit of Bayesian models is their optimisability through machine learning to accelerate the development of ITSs and to refine models based on the learner's data either during the interaction (Schadenberg et al., 2017) or even beforehand (Arroyo et al., 2004; Ferguson et al., 2006). Schadenberg et al. (2017), for instance, based their lesson planning on the likelihood whether the learner will answer correctly or not. This likelihood is modeled with just two parameters, namely, the user's learning ability and the task difficulty. Both are fitted during the interaction while observing the learner's performance to optimize the knowledge tracing and, with that, the lesson planning (Schadenberg et al., 2017).

In addition, Bayesian models are also able to model changes over time, e.g., in Dynamic Bayesian Networks (DBNs). Similar to traditional Bayesian models, DBNs can be used by an ITS to decide what to do next based on the current knowledge about the learning situation. But, they also allow for a more detailed planning of the next steps by simulating the future course of the interaction. Early versions of DBN-like constructs can be found in the Andes physics tutor (Conati, 2002) and Prime Climb math tutor (Conati et al., 2002; Conati and Maclaren, 2009), which model the learner's goals and affective states, while using a non-Bayesian update rule for better scaling. However, probably the most common DBN-like Bayesian model to trace the learner's knowledge is called Bayesian Knowledge Tracing (BKT) (Corbett and Anderson, 1994). It is based on a Hidden Markov Model (HMM) consisting of a latent (the skill-knowledge of just one skill) and an observable variable (the answer correctness) and often serves as a basis for more complex models of knowledge tracing (e.g., de Baker et al., 2008; Lee and Brunskill, 2012; Spaulding et al., 2016). Spaulding et al. (2016), for instance, proposed the Affective BKT model to trace the language reading skills of children in a cHRI. They introduced two further observable variables, namely "smile" and "engagement", to enable the system to take the affective and cognitive state of the child into account and to calculate the skill belief correspondingly. Their evaluation showed that this model outperforms traditional BKT-based models for knowledge tracing in educational settings (Spaulding et al., 2016). Similarly, Käser et al. (2014a) extended the traditional BKT and defined a comprehensive DBN to trace the knowledge of all skills to be learned in just one network. This allows for the system to trace the learner's knowledge about each skill individually and, additionally, to represent and reason about skill interdependencies, which in turn allows to specify the best learning order of all skills or even to let the system search for it autonomously. Their evaluation demonstrated that this more detailed model outperforms the traditional BKT with regard to the accuracy of traced skill beliefs, at least in domains with skills that are interdependent (Käser et al., 2014a). In addition to these examples, many other extensions and variants of the basic BKT can be found that include the item difficulty (Pardos and Heffernan, 2011), forgetting (Khajah et al., 2016), extended learning states (Zhang and Yao, 2018), the time between attempts (Qiu et al., 2011) or investigate the individualization of these models in more detail (Pardos and Heffernan, 2010; Yudelson et al., 2013).

Also generalizations and combinations of Bayesian and logistic models were developed in recent years (Wang et al., 2013; Gonzalez-Brenes et al., 2014; Khajah et al., 2014a,b; Streeter, 2015). Khajah et al. (2014b), for instance, combined Item Response Theory (a logistic model), which allows to model different student abilities and problem difficulties, with a HMM to trace the learner's skill acquisition. Their evaluation showed that each sole model is outperformed by the combination of both types (Khajah et al., 2014b). Streeter (2015), instead, used a more generalized approach called mixture modeling, which also combines both types of models, but shows higher improvements in prediction accuracy on real data (Streeter, 2015). However, the superiority of these more complex hybrid models is mostly only observable when comparing them with fairly simple versions of the basic models. In fact, Zhu et al. (2018) demonstrated that the basic BKT approach extended with temporal information about the performance data already achieves comparable results (Zhu et al., 2018).

5.1.2 Planning the Course of Tutoring Interactions

As already mentioned, the *pedagogical module* of an ITS can rely on the information stored in the *student model*, e.g., the learner's knowledge state, to adapt the tutoring interaction and to plan an optimal path through the curriculum (see Section 3.1). Although many of the reviewed systems above already incorporate basic planning algorithms, they mainly focus on the knowledge tracing part to create an appropriate *student model*. However, the literature also provides information about more elaborated approaches to implement the *pedagogical module*.

A common approach to implement it is to employ Reinforcement Learning (RL) to learn which is the best action to take in each state of a tutoring interaction by simply exploring their effects (Bennane, 2002; Sarma and Ravindran, 2007; Malpani et al., 2011; Bennane, 2013). Sarma and Ravindran (2007), for example, used Q-learning extended with a small answer history to learn which questions or hints to select for the learner in a question-answer game. This scenario is defined as a pattern classification problem, meaning, the student has to classify the pattern (question) by providing an answer (A, B, C or D). This assumption served as a basis to implement artificial neural networks to simulate two types of children, autistic and normally developed, which subsequently are used to evaluate the RL system. The results demonstrate that their system is able to improve the answer success-rate of all learners. Additionally, by considering a small history of the learner's answers as a further information base, it is also able to teach autistic children as effectively as normal learners (Sarma and Ravindran, 2007). Clement et al. (2015), instead, proposed two simple algorithms, which are based on the multi-armed bandit strategy, to plan the next steps of a tutoring interaction. That is, their system either takes the action that maximizes the average learning gain of the student, or explores some new activities, which might be even more beneficial. The whole process is guided by priors gained from expert annotations beforehand, e.g., the impact of actions on students' learning gain and the difficulty of different task types. Although both algorithms share the same priors they differ in the adaptation method and the amount of additional knowledge stored. While one operates without any further information, the second includes a rudimentary memory to store information about the learner's knowledge. Their evaluation shows that even if the ITS does not use information about the learner, it can already lead to a higher learning gain as compared to a human expert's lesson. However, their second algorithm performed even better. Moreover, they concluded that extending their system with a more complex model for tracing the knowledge state of a student might lead to a further improvements regarding the student's learning gains (Clement et al., 2015).

In addition, research on incorporating Partial Observable Markov Decision Processes (POMDPs) for the planning process of an ITS can be found as well (Folsom-Kovarik et al., 2010; Theocharous, 2010; Brunskill and Russell, 2011; Rafferty et al., 2011). While these systems are often able to model the learning interaction fairly accurately, e.g., by extracting the network structure and parameters form a prerecorded dataset (Theocharous, 2010), finding the optimal action policy is often computational unfeasible. However, research on simplifying the planning problem by reducing its complexity is proceeding and approaches either for optimizing the state representation (Folsom-Kovarik et al., 2010) or for online-planning (Brunskill and Russell, 2011; Rafferty et al., 2011) can already be found. The latter just plans a few steps ahead during the interaction (online), instead of inferring a fully specified action policy in advance (offline). Rafferty et al. (2011), for instance, used this method to develop an ITS to teach alphabet arithmetic. They implemented a heuristic forward search that explores usable actions and its effects just two steps beyond the current state. This approach was compared to two different random and one maximum information gain (MIG) policy. The latter just chooses the action that results in the maximum information gain for the learner, if the task is solved correctly. Their results show that the heuristically estimated action-policy achieves a significant faster skill learning than choosing actions randomly. However, compared to the simple MIG algorithm no significant difference is observed anymore, at least for small skill spaces. According to the authors a likely explanation for this finding is that the used knowledge tracing model might have been insufficient (Rafferty et al., 2011).

Finding the optimal teaching policy can also be interpreted as a classification task. That is, the ITS classifies the learner type of a student to associate her with broader groups, for which specific rules are defined, or it classifies specific preferences of the learner, which then can be satisfied by the interaction. Examples for such classifiers are neural networks (Castellano et al., 2007), decisions trees (Cha et al., 2006; McQuiggan et al., 2008) and hybrid methods (Hatzilygeroudis and Prentzas, 2004; Lee, 2007). Hatzilygeroudis and Prentzas (2004), for instance, used a hybrid approach that integrates symbolic rules and neurocomputing called "neurules". They are used to classify the learner type, as well as her preferences, and to define the pedagogical knowledge of an expert system. Both are integrated into an ITS to make pedagogical decisions regarding the interaction course with respect to the inferred information about the learner. In general, those "neurules" are space and time efficient and offer robust inference mechanisms. Furthermore, they can be constructed incrementally and updated easily to refine the expert knowledge base later on. However, they can not represent fussiness, e.g., for modeling students' knowledge, or structural knowledge, e.g., the knowledge stored in a domain model (Hatzilygeroudis and Prentzas, 2004).

Although various of this planning models already incorporate a basic version of a *student model* to keep track of their knowledge and/or abilities, they are not very sophisticated yet, which is an often criticized issue in this line of research. It is often argued that a more effective knowledge tracing method used by the *pedagogical module* of an ITS can be profitable for the learning interaction, e.g., by further increasing students' cognitive learning (e.g., Rafferty et al., 2011; Clement et al., 2015).

5.1.3 Discussion

Reviewing the literature reveals that most published approaches focus either on establishing a *student model* filled with knowledge about the learner's skill mastery and/or engagement, or on the development of sophisticated planning algorithms to choose the next steps to be taken in the tutoring interaction (*pedagogical module*). However, the work on ITSs suggests that both, a sophisticated *student model* and *pedagogical module*, are equally essential for building a tutoring system to provide lessons as beneficial as possible. Because of this, some planning approaches already try to combine both, e.g., by classifying children's learner type to provide well designed default learning content suitable for the respective needs and preferences (e.g., Hatzilygeroudis and Prentzas, 2004; Castellano et al., 2007). But although these models are already rather complex in their planning mechanics, e.g., by basing it on neural networks or even on a hybrid approach that merges symbolic rules and neurocomputing, they just aim for sorting the learner into rather rough groups instead of keeping track of her knowledge to establish an elaborated *student model*.

Also first approaches based on RL can be found that incorporate a rudimentary *student model* (Malpani et al., 2011) or a small history of the learner's answers (Sarma and Ravindran, 2007) to represent or infer her already attained knowledge. In fact, they demonstrated that such fairly simple extensions can already increase the efficiency of an ITS. Moreover, basing the planning on RL allows for learning the pedagogical rules from a dataset or the interaction itself, which was shown to yield good results (Chi et al., 2009). However, these approaches to implement the *pedagogical module* contain also some weaknesses. The resulting models are often inflexible, because they are designed for a specific type of task, e.g., a sequential decision task (e.g., Cakmak and Lopes, 2012), tested only on simulations without the unpredictable noise of real learners (e.g., Sarma and Ravindran, 2007; Malpani et al., 2011) or ignore the uncertainty about learners' real knowledge (e.g., Cakmak and Lopes, 2012). In addition, employing RL for larger learning domains, e.g., language learning, or extending the *student model* by considering also learners' engagement, i.e., affective and cognitive states, enlarges the state space and, thus, strongly increases the time needed to personalize and to learn how to behave until it can be used in the wild effectively. Nevertheless, all these models already demonstrated that basing the planning mechanics of the *pedagogical module* on knowledge about the student, even when this information base is fairly simple, can significantly increase the effectiveness of an ITS.

In general, to establish a more elaborated *student model* two different types of approaches for tracking learners' knowledge are commonly used. First, assumption-free approaches that allow for the model to learn the latent structures within a dataset by itself and, thus, make the time-consuming task of defining them by hand unnecessary. But this increases the difficulty of keeping track of what the system learned and how it will behave in each situation, which, however, is precisely what is often desired in the realm of developing educational software. Experts want to control what and how children should learn, so that it is in line with research in pedagogic and pedagogical psychology. Furthermore, there is a lack of huge datasets to train these models. Consequently, this type of model does not provide the best basis for the development of a SARTS.

Assumption-based models, instead, require predefined assumptions, preferably provided by human experts, which complicates their development, but often maintain their interpretability and understandability. Commonly used assumption-based knowledge tracing approaches build on Bayesian methods, but the questions about the optimal choice between those and logistic models is not fully resolved yet. Although these models already have been compared (Gong et al., 2010), the results are not conclusive, probably do not generalize properly and, thus, are not applicable to each domain. In general, logistic models are favored for memory building processes since the knowledge state is modeled more naturally by gradual changes. In contrast, Bayesian methods, e.g., BKT, often apply a more discrete state transition from not known to known, which allows to model the understanding and sense-making processes in fine-grained knowledge components. Furthermore, the latter often provides a specific graphical structure to model influences of each included attribute. This enables one to understand the likely effects and decisions of such models easily and, thus, allows to prove their validity in the educational context. Additionally, a subset of these approaches allow for a simulation of the learner's development based on the influences of different variables to plan the next tutoring steps. They further seem to be easily extendable and, thus, provide a suitable basis for an incremental implementation of a *student model*. In fact, Khajah et al. (2014a) showed that a BKT can easily be transformed into a hybrid model by including the problem difficulty or general student abilities as latent factors (based on logistic models) directly into the BKT so that they can influence the tracing process explicitly. Their results showed that a BKT model with these rather simple extensions can already outperform the traditional BKT, as well as basic logistic models (Khajah et al., 2014a).

In addition, since BKT is a specific type of DBN, it can easily be extended by a decision component, e.g., to choose the next problem difficulty, leading to a model similar to a Dynamic Bayesian Decision Network (DBDN) (Russell and Norvig, 2010, p. 664ff.) or a POMDP (Russell and Norvig, 2010, p. 658ff.). While solutions of the latter are optimal plans that are computed offline and conditioned on future observations, the former can be regarded as a computational representation of POMDPs, which determines solutions for finite time horizons in an online fashion (cf. Polich and Gmytrasiewicz, 2007). This enables the ITS to select actions and to reason about their effects, so that the learners will receive a tutoring interaction which they most likely benefit from. However, finding an optimal action policy is often computationally intractable. Although research on compressing the state space by means of new representation methods to make them tractable even in larger learning domains has already been done (Folsom-Kovarik et al., 2013), this approach might not be applicable in all domains and, in particular, not in language learning. Another possibility is to use online-planning algorithms, e.g., a heuristic forward search (Brunskill and Russell, 2011; Rafferty et al., 2011). However, this type of algorithm is not guaranteed to explore the whole state and action space to find the optimal solution, since in larger spaces it is limited through a time constraint of just a few seconds to maintain its tractability. But nevertheless, within a domain with a limited or smartly defined state space and a suitable online planning algorithm, POMDPs or DBDNs, respectively, still provide a huge potential to combine information stored in the *student model* with the *pedagogical module* for planning the next steps in a tutoring interaction to optimally address the learners' cognitive learning.

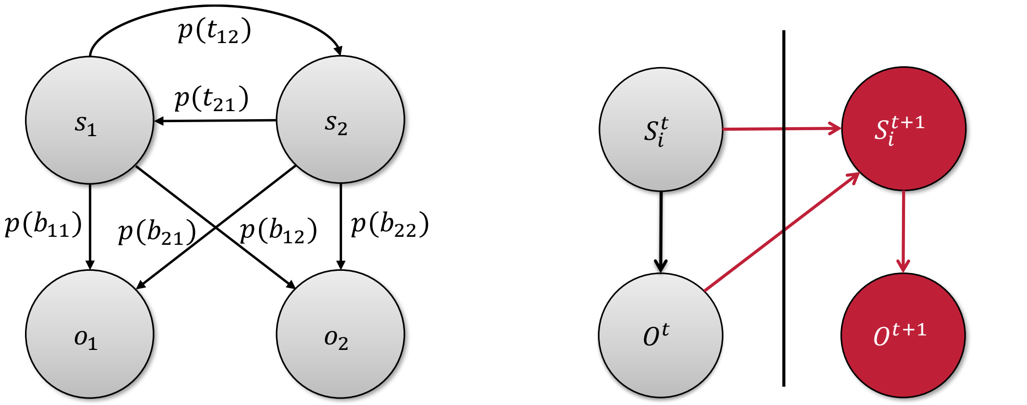
In conclusion, extending the traditional BKT to build a DBDN poses a promising basis for implementing and combining the *student model* and *pedagogical module* of an SARTS. First, the underlying BKT is easily extendable and, hence, can also handle further information about the learner, e.g., the engagement. Second, integrating the decision-making right into the BKT allows for actions to influence the tracing process, which, in fact, was already shown to improve the tracing results (Khajah et al., 2014a). Third, in the chosen domain of word learning of a foreign language, the state and action spaces can be modeled so that the planning process stays tractable. In general, the state portrays the words to be learned whose number, however, can be several tens of thousands. But, the state space can be divided into small "chunks" so that each chunk just includes a small portion of the vocabulary. Letting the planning algorithm just work within the current chunk reduces the costs for planning the next steps dramatically. After a chunk is mastered by the student, the system simply switches to the next chunk of skills and goes on with the teaching process. This is also an established practice in traditional classroom environments, in which teachers also do not teach the whole vocabulary of a language at a time. In addition to the already mentioned benefits, modeling the *student model* and *pedagogical module* tightly coupled also allows to represent some of the interconnections between the different learning dimensions. This is important since the inevitable influences of other dimensions can change the profit for the actions executed by a SARTS (see Section 2.1.6). However, incorporating the influences right into the model enables the tutoring system to simulate different actions and action combinations based on the current information about the learner and to choose them accordingly to optimize the tutoring experience with respect to all dimensions of learning.

5.2 Adaptive Bayesian Knowledge Tracing (A-BKT)

In general, BKT is an implementation of a Hidden Markov Model (see Figure 5.1a), which is defined by the following elements:

- $\mathbb{X} = \{s_1, \ldots, s_n\}$ represents the latent state space
- $\mathbb{O} = \{o_1, \ldots, o_m\}$ is the set of possible observations
- $T \in \mathbb{R}^{n \times n}$ is the transition matrix, where $p(t_{ij})$ is the probability of moving from state s_i to s_j,
- $\Omega \in \mathbb{R}^{n \times m}$ is the observation matrix, where $p(b_{hi}) = p(o_h|s_i)$ is the likelihood of observing o_h in state s_i,
- $\pi \in \mathbb{R}^n$ is the prior distribution, where $\pi_i = p(s_i)$ is the probability that s_i is the initial state for S

The BKT approach depicted in Figure 5.1b adopts this definition so that it contains a latent variable S that represents the skill and an observable variable O that models the user's answer. Although both are classically assumed to be binary, this is already sufficient to represent the skill mastery of the user and to calculate the likelihood of a correct answer given the current skill belief (see Equations 5.1). However, this way of modeling the learner's knowledge requires a separate BKT instance for each single skill.



(a) The HMM is defined by the latent states s_1 and s_2, the possible observations o_1 and o_2, the likelihoods $p(b_{hi})$ to observe an observation o_h in state s_i and the transition $p(t_{ij})$ describing the probability to move from s_i to s_j.

(b) In the traditional BKT the observed answer O^t depends just on the learner's current belief about skill S_i^t (latent). Both, O^t and S_i^t, influence the belief update for S_i^{t+1} in the next time slice.

Figure 5.1: Graphical representation of a traditional HMM (a) and BKT (b).

To update the current belief about a skill being mastered or not, only the observation Ω and transition probabilities T are used to calculate the posterior $p(S_i^{t+1})$. The emission probabilities, which are contained in Ω, are described by the "slip probability" $p(slip)$, which is the likelihood of answering wrongly although knowing the skill, and the "guess probability" $p(guess)$, which is the likelihood answering correctly without knowing the skill. In addition, the transition probability is given by $p(t)$, which represents the skill transition from unknown (s_1) to known (s_2) (cf. Equations 5.2-5.4).

$$p(O^{t+1} = correct) = p(S_i^t) \cdot (1 - p(slip)) + (1 - (S_i^t)) \cdot p(guess) \tag{5.1}$$

$$p(S_i^{t+1}) = p(S_i^{t+1}|O^t) + (1 - p(S_i^{t+1}|O^t)) \cdot p(t) \tag{5.2}$$

where

$$p(S_i^{t+1}|O^t = correct) = \frac{p(S_i^t) \cdot (1 - p(slip))}{p(S_i^t) \cdot (1 - p(slip)) + (1 - p(S_i^t)) \cdot p(guess)} \tag{5.3}$$

$$p(S_i^{t+1}|O^t = wrong) = \frac{p(S_i^t) \cdot p(slip)}{p(S_i^t) \cdot p(s) + (1 - p(S^t)) \cdot (1 - p(guess))} \tag{5.4}$$

Although this basic model can already be used for choosing the next skill to address, e.g., the next vocabulary item with the lowest belief about the learner's mastery level, no information about how the skill can be addressed is represented yet. To achieve this, different options can be taken into account. Khajah et al. (2014a), for example, added the task difficulty with which a skill can be addressed as a latent factor to the basic BKT. This factor directly influences the likelihoods $p(slip)$ and $p(guess)$ so that the likelihood of observing a correct answer given a particular skill belief will be adapted with respect to the task difficulty and, with that, also the tracing process itself (Khajah et al., 2014a). However, this approach does not yet allow to directly choose the next and, in particular, the best task difficulty with regard to the learner's current abilities.

For this purpose it is required to enable the ITS to simulate the effects when providing a specific task with a certain difficulty level and, based on the results, choose the task with the highest learning gain while still not overstraining the learner. Therefore, the Adaptive Bayesian Knowledge Tracing (A-BKT) model proposed here incorporates a decision node A as an extension to the traditional BKT approach, which represents the possible actions to be taken by the system during the tutoring interaction (see Figure 5.2). Moreover, the action decision node is influenced by the learner's knowledge state, which further allows to chose an appropriate action not only based on the likelihood to observe an correct answer but also with respect to the skill mastery. For the moment, the A-BKT actions only describe the different task difficulties, which was already demonstrated to be beneficial for the tracing process by Khajah et al. (2014a). However, later the model's action space can be extended to also address the learner's engagement, as well as affective and perceived learning.

The resulting A-BKT model is comparable to a DBDN, which is defined as follows:

- $\mathbb{X} = \{s_1, \ldots, s_n\}$ represents the latent state space of a skill S_i (e.g., mastered or not mastered)
- $\mathbb{A} = \{a_1, \ldots, a_m\}$ the set of (teaching) actions (e.g., task difficulties)
- $\mathbb{O} = \{o_1, \ldots, o_l\}$ the set of possible observations (e.g., correct or wrong answer)
- $T \in \mathbb{R}^{n \times m \times l \times n}$ the transition matrix, where $p(s_j | s_i, o_y, a_x)$ is the probability of switching from state s_i to s_j when observing o_y after executing a_x
- $\Omega \in \mathbb{R}^{n \times m}$ is the observation matrix, where $p(o_y | s_i, a_x)$ is the likelihood of observing o_y in state s_i if action a_x is applied
- $\pi \in \mathbb{R}^n$ is the initial distribution, where $\pi_k = P(S_i = s_i)$ is the probability that s_i is the initial state of skill S_i
- $U = \{u_1, \ldots, u_n\}$ the set of utilities, where each u_z values a possible consequence of an action or a set of actions

As a first step from BKT towards a DBDN, just the action decision node A is added. Instead of the utilities U, a simple comparison between the current belief state and the desired final belief state for the learning interaction is used. Usually, the utility nodes are used to control the behavior of decision networks by assigning a numerical utility to each possible consequence of an action that represents the respective benefits. This allows to calculate the expected utility and to choose a set of actions to maximize it. However, finding suitable utility values that result in the desired system behavior is a complex task and should be based on an empirical dataset, in the optimal case.

By introducing a separate node for tutoring actions, their influences are modeled as a causal relationship between A and O, meaning, actions directly influence the likelihood of observing a correct answer. Additionally, actions also influence the skill belief update at time $t+1$, which, in consequence, allows for an impact-simulation of tutoring actions beforehand. To keep the model simple, the current action space only consists of three different task difficulties (easy, medium, hard). For example, if the skill belief appears relatively high, which means the skill is nearly mastered, an easy task would provide no challenge and the A-BKT model assumes a high likelihood of observing a correct answer, which in

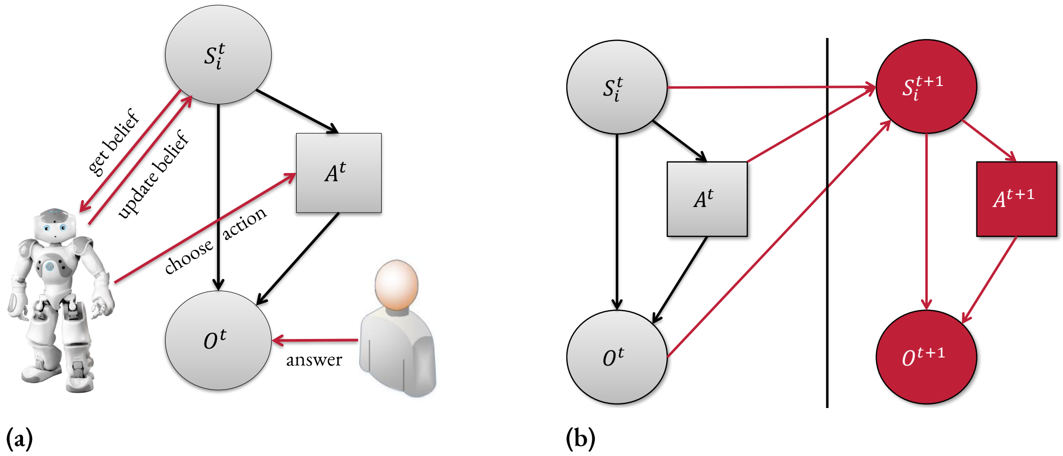


Figure 5.2: Dynamic Bayesian Decision Network based on BKT. (a) The SARTS chooses the next skill S_i^t and action A^t for time step t based on the current belief over all skills/networks (reprinted with permission from Bergmann et al. (2017)). (b) After observing an answer O^t from the learner, this observation, the used action A^t and the previous skill belief of S_i^t are used to update the skill belief for S_i at time $t+1$ (taken and redesigned with permission from Schodde et al. (2017a)).

turn would only result in a relatively minor benefit for training that skill. In contrast, if the skill belief is assumed to be rather low and a hard task is given, the learner would barely be able to solve the task, which results in a low likelihood of observing a correct answer and, thus, in a smaller (or non-existent) learning gain. A task of adequate difficulty, instead, can push the learner into the ZPD, which results in a higher learning gain supported by a feeling of flow (cf. Basawapatna et al., 2013; Craig et al., 2004).

To allow for a more fine-grained decision process the structure of the latent variable S that stores the skill belief is modified. Each variable S_i can now attain six discrete values (states) instead of just two (mastered or not mastered), which corresponds to six bins for the belief state (0%, 20%, 40%, 60%, 80%, 100%). In contrast to the traditional BKT, the discretization into six bins allows to model not only the uncertainty about the current state but also to represent the learners progress in more detail. That is, it allows, for instance, to represent that the system is 80% sure that a learner mastered a skill to 40%. This is required to model the influence of each action on the likelihood of observing a correct answer more accurately. For example, if the skill belief is low (probability mass mainly contained in the first bin), a high task difficulty would result in a low likelihood to observe a correct answer. In contrast, a high skill belief, e.g., probability mass mainly contained in the last two bins, allows for the system to select a difficult task, since the likelihood to observe a correct answer rises to a reasonable level. Consequently, this modification results in an increased flexibility, whereas the complexity that would arise when applying continuous latent variables is avoided.

In addition, the prior distribution π is defined to be uniform so that all states are equally likely, which represents the model's initial uncertainty about the learner's knowledge. However, the observable variable O remains binary as in the classical BKT and still represents whether the learner's answer was correct or not. Although this is a rather simple implementation it is still sufficient to model the learner's answer behavior in the chosen setting, since the learner can either select the correct object associated with the taught word and, thus, answers correctly or choose a wrong object, which would result in a wrong answer. As mentioned in Section 4.3 this is comparable to a CBM in which the learner's final answer just has to satisfy a predefined constraint.

But with these changes the classical BKT update function is not applicable anymore. This function is based on rather simple assumptions about guessing $p(guess)$ and slipping $p(slip)$ in the answer process (see Equation 5.2), which define the complete probability distribution $p(O^t|S_i^t)$. Since S_i^t is split into six possible states instead of being binary and the additional action decision node A^t influences the observable O^t as well, the resulting update function has to be adapted to consider these changes. To resolve this, a standard Bayesian update is applied for all $s_k \in S_i^{t+1}$ to calculate the next skill beliefs with respect to $p(o^t|S_i^t, a^t)$ and the transition probability $p(s_k|S_i^t, o^t, a^t)$.

$$p(s_k) := p(s_k|o^t, a^t) \tag{5.5}$$

$$= \sum_{s_j \in S_i^t} [p(s_j|o^t, a^t) \cdot p(s_k|s_j, o^t, a^t)] \tag{5.6}$$

$$= \sum_{s_j \in S_i^t} [\frac{p(o^t|s_j, a^t) \cdot p(a^t|s_j) \cdot p(s_j)}{p(o^t, a^t)} \cdot p(s_k|s_j, o^t, a^t)] \tag{5.7}$$

Based on this, the model can now be used to decide which skill would be best to address next and which action to choose to address this particular skill.

5.2.1 Predictive Decision-Making

The developed algorithm for skill selection is comparable to the vocabulary learning technique called *spaced repetition*, which is implemented, for instance, in the Leitner system (Leitner, 1972, p. 64ff.). It enables the system to choose the next skill to address that maximizes the beliefs of all skills while balancing the single skill beliefs among each other. To achieve this, skills with lower beliefs are prioritized and repeated more frequently while well known skills are addressed with a lower frequency (*spaced repetition*). This behavior can be modeled easily by calculating the Kullback-Leibner divergence (KLD) between the current and the desired skill belief $p(S_{opt})$. Since the application of the KLD requires at least some probability mass in each bin, $p(S_{opt})$ is defined by containing $\approx 99.999\%$ of the probability mass in the last bin, meaning, being $\approx 99.999\%$ sure that the learner has mastered a skill to 100%.

$$next\_skill = \underset{S_i^t \in \mathbb{S}}{\operatorname{argmin}} \left[\alpha(S_i^t) \cdot KLD(p(S_i^t), p(S_{opt}))\right] \tag{5.8}$$

$\mathbb{S}$ represents the set of all addressable skills S_i that can be taught. As mentioned above, in the scope of language learning this set includes a huge number of skills/words and, thus, should be split into smaller chunks, so that the skill selection algorithm has to evaluate just a small part of the possible skill space and remains tractable. To regulate the skill occurrence frequency and to achieve the proposed *spaced repetition* effect of maximizing, as well as balancing all skills, the factor $\alpha(S_i^t)$ is added. It ranges from 0.0 to 1.0 and is decreased by 0.3 each time a specific skill S_i is addressed and is increased again by

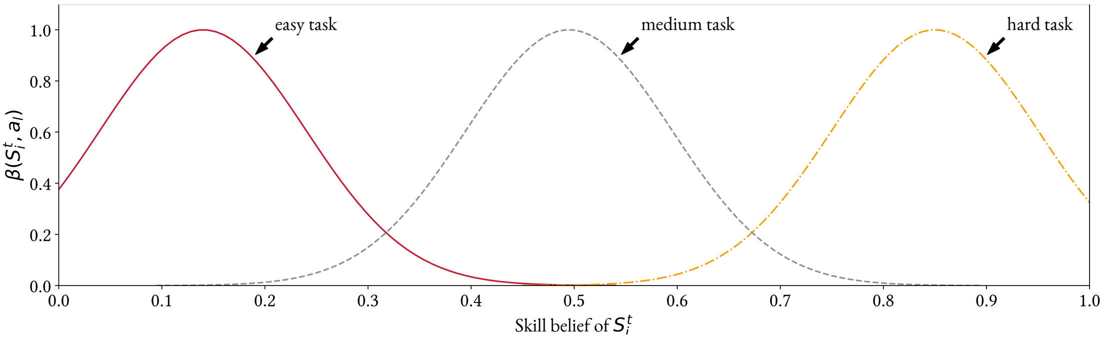


Figure 5.3: Gaussian distributions with respect to each task difficulty used for $\beta(S_i^t, a_l)$.

0.15 if another skill is being practiced. This prevents the system from focusing on just one specific skill by addressing it continuously until a sufficient number of correct answers is provided, the skill belief is high enough and in balance with the remaining skills again. However, this results in a slightly different behavior as in the traditional *spaced repetition* system.

In the *spaced repetition* system all words are practiced at least once until the "spaced" repetition phase starts, in which words with a higher error frequency are repeated more often than well known words. The current implementation of the A-BKT model, in contrast, starts to practice some few skills first. But if the provided answers contain a lot of mistakes, the algorithm focuses on this smaller set of weak skills before it introduces new skills. This ensures that the learner has already a good basis of the first skills instead of getting overwhelmed by too much content.

After a skill is selected the system has to decide with which tutoring action it should be addressed. As mentioned before, the task difficulties are considered as tutoring actions, which represent abstract tasks that are mapped onto concrete exercises or pedagogical acts in the SARTS architecture later on. Similar to the skill selection, the action selection is modeled as a minimization problem. That is, the system chooses the *next_action* that minimizes the difference between the predicted and the desired skill belief distribution to increase the learner's knowledge gain for the selected skill.

$$next\_action = \underset{a_l \in A^t}{\text{argmin}} \left[\beta(S_i^t, a_l) \cdot KLD(p(S_i^{t+1}|a_l), p(S_{opt}))\right] \tag{5.9}$$

where

$$p(S_i^{t+1}|a_l) := \sum_{o_m \in O^t} \sum_{s_j \in S_i^t} \left[p(s_j|o_m, a_l) \cdot p(s_k|s_j, o_m, a_l)\right], \forall s_k \in S_i^{t+1} \tag{5.10}$$

$$= \sum_{o_m \in O^t} \sum_{s_j \in S_i^t} \left[\frac{p(o_m|s_j, a_l) \cdot p(a_l|s_j) \cdot p(s_j)}{p(o_m, a_l)} \cdot p(s_k|s_j, o_m, a_l)\right], \forall s_k \in S_i^{t+1} \tag{5.11}$$

Here, $p(S_i^{t+1}|a_l)$ is used to predict the effect of applying the current action a_l to skill S_i by incorporating the KLD to compare the skill belief with the desired skill belief $p(S_{opt})$. This procedure also substitutes the process of calculating and maximizing the expected utility for now and, hence, avoids the definition of appropriate utility values for each state. As a further simplification and as a control mechanism to refine the process of selecting the "best" action for each state, the factor $\beta(S_i^t, a_l)$ is introduced. It is based on a series of overlapping Gaussian curves distributed over the full belief space of a skill S_i, while each curve defines the field of application for a specific action a_l (task difficulty, see Figure 5.3). Consequently, $\beta(S_i^t, a_l)$ modifies the KLD so that it is higher if an action is selected, which is assumed to be inappropriate for the current skill belief (e.g., high task difficulty for low skill mastery). However, if an action is assumed to perfectly fit the current skill belief about the learner's skill mastery, $\beta(S_i^t, a_l)$ does not affect the KLD.

Overall, the presented definition of the A-BKT approach results in a model that selects an easy task if the skill mastery is believed to be low, a hard task if it is high, and medium in-between. This is the another difference to simple *spaced repetition* systems, since the difficulty of tasks does not change within those systems. However, the goal of the developed adaptation strategy is to create a feeling of flow, which can lead to better learning results (Craig et al., 2004; Hamari et al., 2016). Consequently, it strives not to overburden the learner with too difficult tasks or to bore them with too easy tasks, both of which may lead to frustration and, with that, hamper learning (Engeser and Rheinberg, 2008; Habgood and Ainsworth, 2011).

5.3 Study 1: Evaluation with Adults

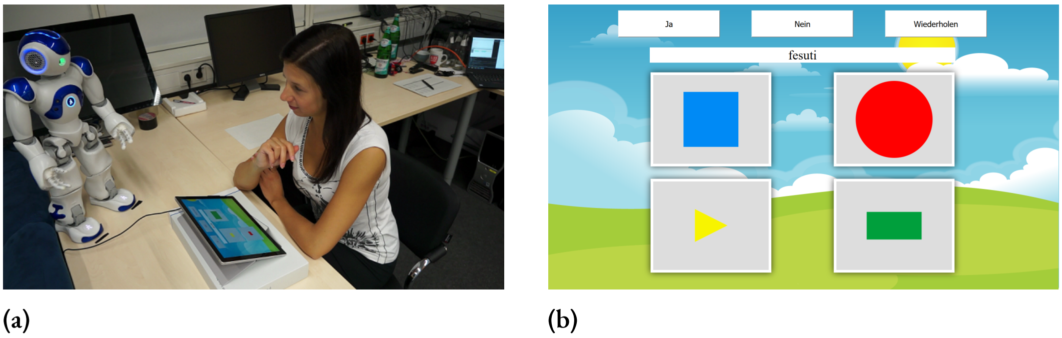
To study the efficiency of the proposed A-BKT model on vocabulary learning of a foreign language, an evaluation study is set up. As a simplification, this first evaluation is conducted with adult learners. Further, it is based on the "I spy with my little eye ..." game implemented in the SARTS described in Chapter 4, which is inspired by the previously identified characteristics of a typical tutoring interaction in German kindergartens. The major objective of this study is to evaluate the effects of the A-BKT model on learners' word learning performance (cognitive learning). Further, it is hypothesized that:

- **H1:** Participants who learn with and from the adaptive system using the A-BKT model will perform better during training than those who learn without a personalized tutoring strategy.
- **H2:** Participants who learn with and from the adaptive system using the A-BKT model will perform better in the post-test than those who learn without a personalized tutoring strategy.

5.3.1 Study Design

The study uses a between-subjects design with two conditions, adaptive and random, where the type of training varied between them. In the control condition all skills are taught with a medium task difficulty in a randomized order for 30 rounds (each word three times), whereas in the adaptive condition skills and actions are chosen by the A-BKT model as explained in Section 5.2.1. However, if the learner makes too many mistakes, the A-BKT model focuses on a small set of weakest words first, until teaching the remaining skills. In order to avoid ignoring some words, until the 30 rounds are reached, the SARTS is forced to teach each skill at least once.

As described above, within the scope of language learning the skills relate to the foreign language vocabulary (target words) and an action refers to a specific task difficulty used in the game to address a



(a) (b)

Figure 5.4: Experimental setup (a) with a participant sitting in front of a tablet displaying the graphical user interface (b). The robot Nao stands next to the tablet slightly rotated towards the user (both reprinted with permission from Schodde et al. (2017a)).

skill. In the "I spy with my little eye ..." game, the task difficulty is manipulated by using one to three distractors that are shown together with the target object on the screen. For instance, an easy task only includes one distractor object, whereas a hard task has three distractors (see Figure 5.4b for an example of a hard task). The distractors are chosen with respect to the skill beliefs the system has of the learner so that the set of objects mainly consist of items for which the Vimmi word is still/mostly unknown. This is intended to maintain the task difficulties as long as possible and to prevent the learner from guessing correctly supported by the process of elimination. This is especially important, since a better learning performance is expected when the learner has to expend the right amount of cognitive effort, i.e., not too hard nor too easy tasks, because this can support a feeling of flow (cf. Basawapatna et al., 2013; Craig et al., 2004). However, at a certain point the learner will know too many words or skills, respectively, so that finding a distractor set that cannot be thinned out by exclusion becomes impossible.

Since this study is conducted with adults, who already have appropriate reading skills, the GUI is extended with three additional buttons to allow for a fully autonomous interaction as of the start. This enables the user to provide further input in terms of "yes" and "no" answers, as well as to repeat the robot's latest sentence (see Figure 5.4). Alternatively, a sophisticated ASR system could have been used, however, the reliability of buttons is still higher, also for adult participants.

The training materials comprises German–Vimmi word pairs to describe geometrical forms on the tablet screen (see Figure 5.4). Vimmi is an artificial language specifically developed for experimental purposes and its vocabulary is created with respect to Italian phonotactic rules (cf. Macedonia et al., 2010). The goal of Vimmi is to avoid associations with other known words or languages, which could influence the study's results. Especially when working with adult participants there is a huge chance that they have multilingual skills, which can cover a broad variety of languages. Consequently, in the scope of this study ten Vimmi words are used that include four color terms, four shape-encoding terms and two terms describing the size (see Table 5.1).

To assess the participants' performance two different measures are used. First, the learners' response behavior during the interaction is tracked to examine their progress of learning. Second, a post-test on the taught vocabulary in the form of two translation tests, German-to-Vimmi and Vimmi-to-German, is conducted to assess the participants' knowledge state subsequent to the tutoring interaction.

N	German	Vimmi	English translation
1	blau	bati	blue
2	grün	uteli	green
3	gelb	dirube	yellow
4	rot	fesuti	red
5	rund	beropuga	round
6	dreieckig	pewo	triangular
7	quadratisch	tanedila	square
8	rechteckig	paltra	rectangular
9	klein	kiale	small
10	groß	ilado	big

Table 5.1: The ten vocabulary words to be learned with its translation in German and English (reprinted with permission from Schodde et al. (2017a)).

The whole tutoring interaction, as well as the vocabulary post-test afterwards, is recorded with an external camera and the system's decisions in each round (random and adaptive), as well as the corresponding probability distributions for all skills (adaptive only) are logged to a text file for later analysis.

5.3.2 Participants

A total of 40 participants, 20 per condition, with an average age of 24.13 years ($SD = 3.82$ years) took part in this study including 16 males and 24 females. All participants were fluent speakers of the German language and had normal or corrected sight. Further, they had no prior knowledge in the chosen target language and were either paid or received ECTS credits for their participation.

5.3.3 Procedure

Upon entering the lab the participant is randomly assigned to one of the two experimental conditions, either receiving the stimulus in an A-BKT adapted order (adaptive condition) or in a random order (control condition). She is informed that she takes part in an experiment on foreign language learning and is asked to sign an informed consent form. Further, she fills out a questionnaire that covers personal information such as age, nationality and a personal estimation of her general abilities to learn languages and to memorize vocabulary of a foreign language.

Following, a list of the ten Vimmi words is presented to the participant for exactly 30 seconds. This should allow for her to build up a first rough concept (*fast mapping*, see Section 2.2) and to practice the items right from the beginning of the actual game, instead of guessing the meaning of completely unknown words. Afterwards, the learning interaction with the SARTS begins. As mentioned above, it is based on the previously identified structure appropriate for children (see Section 4.1.3.1). After introducing itself, the robot explains the "I spy with my little eye ..." game and starts a test run with the participant. Once this test run is finished and the participant agreed that she has understood the game, the main learning interaction begins. It consists of a total of 30 trials, whereas each trial addresses just one vocabulary item. That is, the SAR describes one of the objects displayed on the tablet screen

	Adaptive (A)		Control (C)		A, C	
	M	SD	M	SD	M	SD
F7	3.75	1.37	4.00	1.17	3.88	1.27
L7	6.90	0.31	5.15	1.69	6.03	1.49
F7, L7	5.33	0.69	4.58	1.12		

Table 5.2: Means (M) and standard deviations (SD) of correct answers for the initial quarter of the training interaction (first seven items – F7) and the final quarter (last seven items – L7) in each condition, as well as the inter-model (A, C) and intra-model (F7, L7) means and standard deviations (reprinted with permission from Schodde et al. (2017a)).

with the major part of the description being in German, except for the target word to be learned. An exemplary task can look as follows: *"Ich sehe was, was du nicht siehst und das ist **bati**"* ("I spy with my little eye, something that is **bati**"). After 30 trials the game is finished and the Nao robot thanks the participant and says goodbye.

Subsequent to the interaction with the SARTS, the participant's learning performance is assessed with a post-test, which is conducted in an interview with the experimenter. The participant has to translate the ten vocabulary items from German to Vimmi and, likewise, from Vimmi to German in a randomized order. Finally, the experimenter thanks the participant and guides her out of the room.

5.3.4 Results

The following section summarizes the evaluation results of the A-BKT model with adult learners with respect to their performance during the training stage of the game (H1) and their performance in the post-test afterwards (H2).

5.3.4.1 Learning Progress During Training

In order to assess learners' progress during training the number of correct answers provided in the initial quarter of the tutoring game (first seven items) is compared to the corresponding number in the final quarter (last seven items). When an item occurred repeatedly within the initial quarter the first occurrence is taken into account. Similarly, when an item occurred repeatedly within the final quarter just the last occurrence is considered. In both cases, the quarter is expanded so that exactly seven distinct items are included.

A two-way ANOVA with training phase (initial, final) as a within-subjects factor and training type (adaptive, control) as between-subjects factor was conducted. The results are summarized in Table 5.2 and Figure 5.5. Not surprisingly, there is a main effect of training phase at a significant level ($F(1, 38) = 66.85, p < .001, \eta^2 = .64$), showing that learners' performance was significantly better in the final phase as compared to the initial phase. Participants achieved an average of 3.88 ($SD = 1.27$) correct answers in the first quarter of training, as opposed to an average of 6.03 ($SD = 1.49$) items correctly selected in the final quarter. More interestingly, there is also a main effect of training type ($F(1, 38) = 6.52, p = .02, \eta^2 = .15$), which indicates that participants who learned in the adaptive condition had a

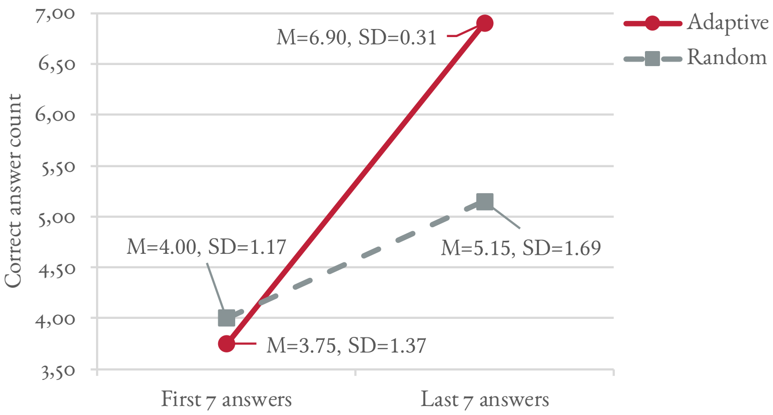


Figure 5.5: Mean numbers of correct answers at the beginning (first 7) and end (last 7) of the interaction in the different conditions (taken and redesigned with permission from Schodde et al. (2017a)).

significant higher average score of correct answers ($M = 5.33$, $SD = .69$) as compared to learners in the control condition with an average of $M = 4.58$ ($SD = 1.12$) correct answers. Finally, the interaction between training phase and training type is also significant ($F(1, 38) = 14.46$, $p = .001$, $\eta^2 = .28$) indicating that the benefit of A-BKT-based training develops over time (see Figure 5.5). While participants' response behavior in the first quarter of training was similar across conditions, a benefit of training with the A-BKT model becomes evident in the final quarter. Participants in the A-BKT model condition achieved an average of $M = 6.90$ ($SD = .31$) correct answers for the last seven distinct skills, whereas participants in the control condition achieved only an average of $M = 5.15$ ($SD = 1.69$) correct answers. In summary, the found results fully support hypothesis H1 that participants will perform better during training when they learn with the adaptive system based on the A-BKT model.

5.3.4.2 Post-Test

Afterwards, the participants' vocabulary learning performance, assessed subsequent to the learning interaction, is analyzed, which was measured with two different translations tests. Paired-sample t-tests were conducted to compare the number of correctly recalled words after the training with the A-BKT model, to that resulting from the randomized teaching strategy in the control condition. For the German-to-Vimmi translation no significant effect could be observed. The participants who learned with the A-BKT model, correctly recalled an average of 3.95 ($SD = 2.56$) words out of ten, while participants in the control condition correctly recalled an average of 3.35 ($SD = 1.98$) words. Likewise, no significant effect for the Vimmi-to-German translation task could be found. Participants' performance after learning with the adaptive model amounted to an average of 7.05 ($SD = 2.56$) correct items compared to participants' performance in the control condition with an average of 6.85 ($SD = 2.48$) correct answers.

Even though, no main effect of training type emerged in the post-test, some details might be worth mentioning. As depicted in Figure 5.6 in the German-to-Vimmi post-test a maximum of ten correct answers was achieved by some participants in the A-BKT model condition, whereas the maximum in

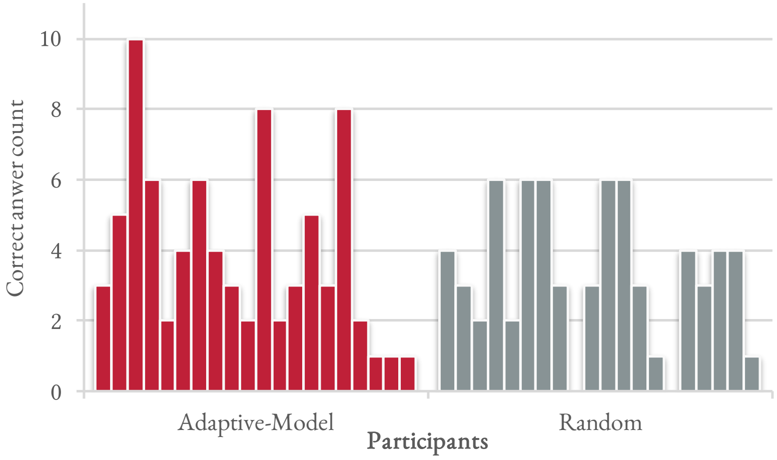


Figure 5.6: Participant-wise amount of correct answers grouped by the different conditions for the German-to-Vimmi post-test (taken and redesigned with permission from Schodde et al. (2017a)).

the control condition amounted to only six correct answers. Moreover, two participants in the control condition did not manage to perform any German-to-Vimmi translation correctly. In the A-BKT model condition, however, all participants achieved at least one correct answer.

To sum up, the observed results do not support the hypothesis H2 that participants who learn with the adaptive system based on the A-BKT model will perform better in the post-test subsequent to the learning interaction.

5.3.5 Discussion

The major goal of the presented study was to evaluate the proposed A-BKT model within a language learning interaction. Therefore, 40 participants were invited to play a language learning game together with the robot Nao, supported by the underlying SARTS. They were either playing in an adaptive condition, in which the A-BKT model was used to select the next skill and the corresponding task difficulty, or in a randomized control condition.

The results show that participants' within the adaptive condition perform better during the training stage of the game, as compared to the control group. Analyzing their response behavior shows not only that participants were able to learn Vimmi words throughout the interaction but also, and more importantly, that they learned more successfully with the adaptive model in comparison to randomized tutoring. This is due to the A-BKT model's ability to prioritize unknown or hard to remember skills, which results in repeated trials for these particular words until the system's belief state becomes similar to those of known words. This strategy outperformed the tutoring of the same material but with equal number of repetitions (three per word) and in randomized order.

In the post-test, however, no significant difference between experimental conditions were found, although a trend towards better performance in the adaptive model condition compared to the control condition could be observed. Different explanations may account for this inconsistent finding. First, the way responses were requested from the learner was not identical in the training sessions and the post-test. During training, pictures reflecting the meaning of the target words were shown, whereas in the post-test the participants merely received a linguistic cue from the experimenter in the form of a word they had to translate. Consequently, the training with the SARTS might have led to stronger associations between linguistic and figurative materials, in particular for words that were difficult to remember since they were repeatedly presented in the adaptive condition. This might have triggered a stronger decline of correct answers for participants who trained with the adaptive model as opposed to those in the control condition. Second, the post-test results measured immediately after the training session could be governed by the strong inter-individual differences among learners, e.g., in the time they need to internalize the knowledge. Consulting the literature revealed other studies that try to cope with this problem by introducing a second and delayed post-test, further called as retention-test, which can be conducted a few days or even weeks later (e.g., Khoshsima et al., 2015; Singer and Gerrits, 2015). In fact, their results demonstrate that significant differences can still appear after a longer period of time. Consequently, introducing a retention-test might reveal results that match learners' performance during the training.

In summary, the A-BKT model revealed promising results for its application in a SARTS within a language learning interaction. Although the post-test results were not entirely conclusive, the A-BKT model demonstrated its ability to support the participants by increasing their learning performance during training and, thus, can be evaluated with children in the next step. However, the study design requires some small but important changes beforehand. First, the post-test style has to be adapted to be similar to the training session and, second, a retention-test needs to be introduced. Finally, the tutoring game, as well as the robot's dialog, have to be adjusted to be suitable and understandable for young kindergarten children.

5.4 Study 2: Evaluation with Children

The initial evaluation of the A-BKT model with adults already revealed promising results and, thus, a second study is conducted to evaluate whether the found effects and benefits can also be attained with young children. To maintain the comparability regarding already found effects as far as possible, the same study setup is used (see Section 5.3.1), which, however, required some small adjustments to be more suitable for young children (see Section 5.4.1). In addition to changing the target group, the assumed weaknesses of the previously applied post-test are resolved to examine whether the post-test results will now reflect the previously found benefits of the A-BKT model. Moreover, the study is conducted in cooperation with partners from the L2TOR project and, therefore, further study conditions are introduced for studying whether the application of iconic gestures can support children's learning performance, in particular when combined with an adaptive system. In summary, this study aims for investigating the following hypotheses:

H1: Children's performance during training is better when target words are accompanied by iconic gestures, than in the case of not using gestures.

N	Dutch	English
1	kip	chicken
2	vogel	bird
3	lieveheersbeestje	ladybug
4	paard	horse
5	aap	monkey
6	nijlpaarden	hippo

Table 5.3: The six English target words to be learned in the evaluation study with its corresponding translation in Dutch (reprinted with permission from de Wit et al. (2018)).

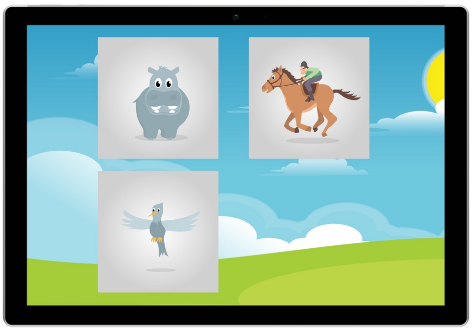


Figure 5.7: Conceptional image of the user interface during the tutoring interaction, showing the image corresponding to the current target word and two distractors (reprinted with permission from de Wit et al. (2018)).

H2: Children show a greater learning gain in the post-tests when target words are accompanied by iconic gestures during training, than in the case of not using gestures.

H3: Children who learn with the adaptive system using the A-BKT model perform better during training as compared to those who learn the target words in randomized order.

H4: Children who learn with the adaptive system using the A-BKT model perform better in the post tests as compared to those who learn the target words in randomized order.

In this section mainly the results with respect to the effects of the adaptive system are discussed, however, all results are reported.

5.4.1 Study Design

To investigate the hypotheses a two (adaptive versus random) by two (gestures versus no gestures) between-subjects design is chosen. While the general basis of the tutoring game remains the same as in the previous study (see Section 5.3.1), it is slightly modified and designed to be more child friendly. First of all, the previously used target words are exchanged with the names of commonly known animals (see Table 5.3) for which matching images are designed to be presented on the tablet screen (see Figure 5.7). Although existing literature defines the optimal number of target words for young children in a short tutoring interaction (e.g., 8-12 words within 60 minutes; Gairns et al., 1986, p. 66) and highlights the importance of repetitions to improve their learning gains (Leung, 1992; Marilyn, 1994; Sénéchal, 1997), there is literally no literature that defines the number of repetitions needed for young children to internalize a word. To ensure a sufficient number of repetitions, while trying not to overstrain the children with an interaction that lasts too long (cf. David Cornish et al., 2009, p. 72f.) or too much learning content (cf. Christ and Wang, 2012), the number of target words is reduced from ten (in the study with adults) to six. Additionally, the target language is switched to English, since a broad repertoire in foreign languages was not expected in kindergarten children. This allows for equipping the children with foreign language skills they can benefit from in their lives later on. However, some children still might

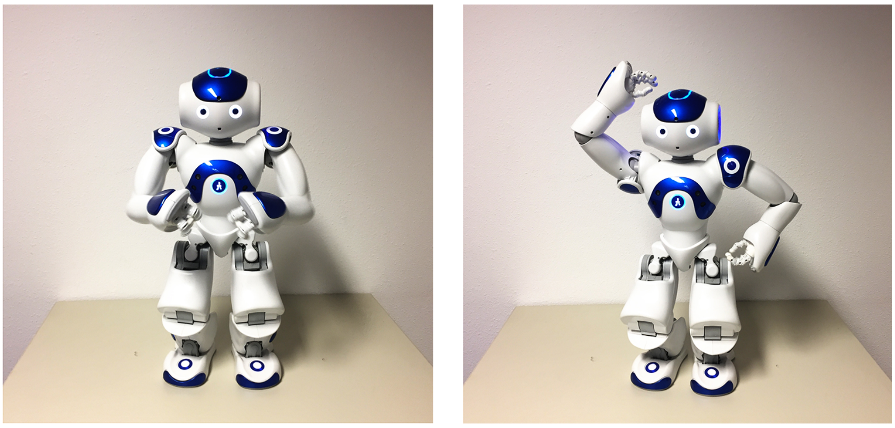


Figure 5.8: Examples of the robot showing iconic gestures for two animals used in the study. Left: imitating riding a horse by grasping imaginary reins and moving the arms up and down; Right: imitating a monkey by scratching armpit and head (reprinted with permission from de Wit et al. (2017)).

have picked up single words in there every day life, e.g., through television shows. To assess this possible prior knowledge a pre-test is introduced, which allows to calculate and analyze the children's relative learning gain by comparing pre- and post-test results. Moreover, a retention-test is introduced to also investigate whether the children need more time to internalize the learned content (see Section 5.3.5). While the pre- and post-tests are adjusted to use the same style of requesting responses as during the tutoring interaction, they are also designed to be clearly distinct with respect to the physical context (laptop versus tablet), the applied voice for spoken content and some characteristics of the used images. Theses changes should ensure to attain more reliable test results from the participants.

In the gesture condition every target word pronounced in English, except during the feedback, is accompanied by a matching iconic gesture (see Figure 5.8 for an example). It is synchronized with the spoken target word so that it is in line with the gesture's stroke. To verify the understandability of these gestures, a perception study was conducted beforehand (de Wit et al., 2017). In this study video recordings of all six gestures performed by the robot were shown to 14 participants, who were asked to indicate which of the six possible target words matches each recording best. The results demonstrated that each developed gesture was sufficiently unique to distinguish between all six target words.

Again, the whole tutoring interaction, as well as the pre-, post- and retention-test, are recorded with an external camera. Further, the system's decision in each round (random and adaptive) and the corresponding probability distributions for all skills (adaptive only) are logged to a text file for later analysis.

5.4.2 Participants

A total of 99 children, recruited from primary schools in the Netherlands, participated in the study. Although all children were allowed to participate, $n = 1$ child had to be excluded because of a missing retention-test result, while $n = 37$ children were excluded, because they are not native Dutch speakers. The final group comprises $n = 61$ children consisting of 29 boys and 32 girls. Their average age was 5.17 years ($SD = 7$ months) and they all had normal or corrected sight.

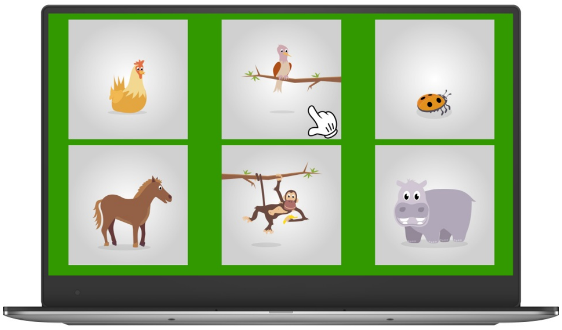


Figure 5.9: Conceptional picture of the pre- and post-tests on a laptop, using a recorded voice and a slightly modified set of animal images (reprinted with permission form de Wit et al. (2017)).

5.4.3 Procedure

A few weeks before the study, children's parents have received information about the goal, setting and course of the study in the primary schools. They are informed that each child needs a completed informed consent form to be allowed to participate in the study and are encouraged to contact the experimenters in case of any ambiguities or uncertainties.

A few days before the study the children are introduced to the robot during a group introduction. As mentioned in Section 4.3, this is intended to lower children's anxiety in the subsequent one-to-one interaction with the robot and is inspired by the work of Vogt et al. (2017) and Fridin (2014a). During the introduction a small background story is told and the differences and similarities between the robot and humans are highlighted. Afterwards, the children (and sometimes teachers) have the chance to shake the robot's hand and to perform dances together with the robot before putting it to bed.

On the study day, each child is randomly assigned to one of the following four study conditions:

1. with random tutoring strategy and no gestures ($n = 16$)
2. with random tutoring strategy and gestures ($n = 14$)
3. with adaptive tutoring strategy and no gestures ($n = 15$)
4. with adaptive tutoring strategy and gestures ($n = 16$)

After the child entered the room, first, the pre-test is conducted to assess her prior knowledge of the six animals names in Dutch and English. During the test she sees images of all six animals on a laptop screen in randomized order, while a recording of a (bilingual) native speaker pronouncing one of the six animal names is successively presented for each word. After the recording has finished, the child is asked to select the corresponding image on the screen (see Figure 5.9). Subsequent to the pre-test, all six animal names are presented with the respective image once in English. During this process each animal is shown in the middle of a laptop screen, while a recording pronounced by a (bilingual) native speaker was played, e.g., *"Look, this is a **chicken**. Do you see the **chicken**? Click on the **chicken**!"*, with the whole sentence being verbalized in Dutch except for the target word, which is presented in English. This procedure should ensure that each child has the chance to establish a first rough concept with the correct word mapping (*fast mapping*) and is further intended to correct early misconceptions established during the pre-test. Moreover, this procedure should prevent the child from randomly guessing the answers in the first rounds of the actual tutoring game.

Afterwards, the tutoring interaction starts in which the robot and the child play 30 rounds of the "I spy with my little eye ..." game. After the robot explained the game, a test run is started to check the children's understanding. Additionally, the child is asked to press either a green (agreement) or a red smiley (disagreement) as a further validation of whether she understood the game-rules. By pressing the red smiley the interaction is paused and the experimenter steps in to further explain the game procedure. Instead, if the child confirms her understanding the game interaction starts.

Each trial of the tutoring game addresses just one vocabulary item. That is, the robot asks for one of the animals displayed on the tablet screen by saying *"I spy with my little eye something that is a* ***[target word]****"*, with the question again being in Dutch, except for the target word, which is presented in English. In both adaptive conditions the number of distractors and the order of skills are again chosen by the A-BKT model. Further, it is again ensured that each skill is taught at least once. In both random conditions, instead, each skill is taught an equal number of times (each word 5 times) in randomized order, whereas the game difficulty is fixed to medium with just two distractors. After 30 trials, the game is finished, and the Nao robot thanks the participant and says goodbye. Afterwards, the child is asked to fill out a post-test, which is identical to the pre-test. Additionally, this test is repeated at least one week after the experiment to measure the long-term retention of the learned words, too.

5.4.4 Analysis

To analyze the results of both post-tests, also children's prior knowledge is considered. To this end, children's relative learning is calculated as the difference between the pre- and post-tests. Hence, the immediate learning gain describes the difference between the number of correct answers in the pre- and post-test, whereas the long-term learning gain is defined as the difference between pre- and retention-test. In addition, the knowledge decay is calculated as the difference between post- and retention-test. Since each target word is asked only once in English, the test scores are always between 0 and 6.

Also, for comparing children's performance during the interaction the recorded data were preprocessed. Since the presented study incorporated less than 7 target words, which would have been necessary for the previously used first and last quarter analysis (see Section 5.3.4.1), a new method had to be developed. One option could have been to reduce the number of rounds included in the analysis, but this would have resulted in analyzing an even smaller portion of the interaction. To be more flexible and independent of the number of target words a normalized performance score was developed. This method can be used to analyze learners' performance for the whole interaction while being able to take the changing task difficulties into account. Since the varying difficulty results in more or less distractors shown on the tablet, the chance of guessing the right answer is also varying. The here developed normalized performance score maps binary task success (1: correct answer; 0: incorrect answer) onto the span between 0.0 and 1.0. For the most difficult task with five distractors (only in pre- and post-tests) a correct answer receives a score of 1.0, whereas for each missing distractor a value of 0.2 is subtracted.

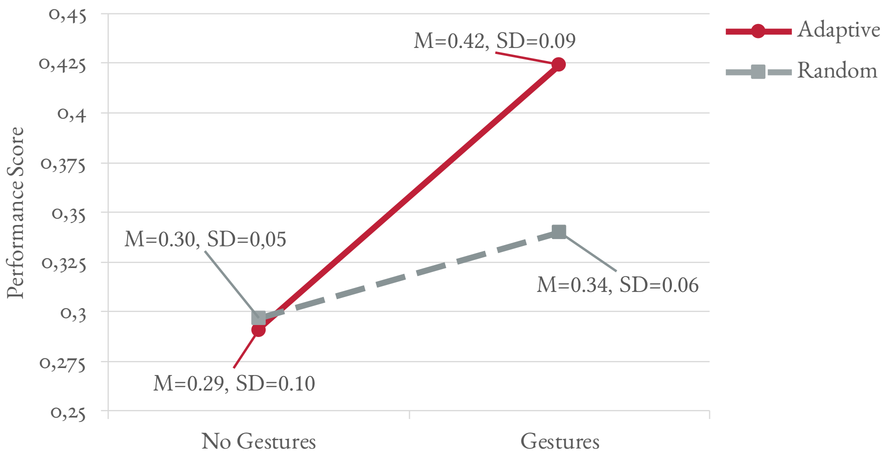


Figure 5.10: Visualization of the learners' performance during training with respect to the different study conditions (taken and redesigned with permission from de Wit et al. (2018)).

This, for example, results in a score of 0.6 for a correct answer in a hard task with three distractors. Note that this is the highest difficulty the children are confronted with during the game, which results in a maximal achievable score of 0.6 on average during training. If a task is answered wrongly, instead, a score of 0.0 is given. Finally, the total training success is accumulated and divided by the total amount of rounds played (30) resulting in a normalized performance score for the whole interaction.

5.4.5 Results

The tutoring interaction lasted 18:38 minutes ($SD = 3{:}03$ minutes) on average. When including the introduction, pre- and post-test, the total session length rises to roughly 30 minutes. To verify that children were generally able to learn new words regardless of the assigned condition the immediate learning gain was analyzed by means of a paired-sample t-test, which shows a significant difference between pre- ($M = 1.75, SD = 1.14$) and post-test ($M = 2.85, SD = 1.61; t(60) = 5.23, p < .001$). Running the same analysis for the long-term learning gain between the pre- and retention-test ($M = 3.02, SD = 1.40$) also reveals a significant difference ($t(60) = 6.81, p < .001$). Correspondingly, the children were able to learn a significant number of words during the interaction and memorize them at least for a few weeks. The analysis of children's knowledge decay, instead, revealed no significant differences indicating that the time between both tests did not result in forgetting or a stronger internalization.

To investigate whether the different study conditions had an effect on the training of skills a two-way ANOVA was carried out. It utilizes the tutoring strategies (adaptive versus random) and gesture use (gestures versus no gestures) as between-subjects and performance during the training as within-subjects factor (see Figure 5.10). As described above, the used performance scores are weighted by the number of distractors and normalized by the number of rounds played. The results show a main effect of gesture use ($F(1, 57) = 18.23, p < .001, \eta^2 = .22$), showing that training with gestures led to a significant higher performance score ($M = .38, SD = .09$) as compared to a learning interaction without gestural support ($M = .29, SD = .08$). Although children in the adaptive condition

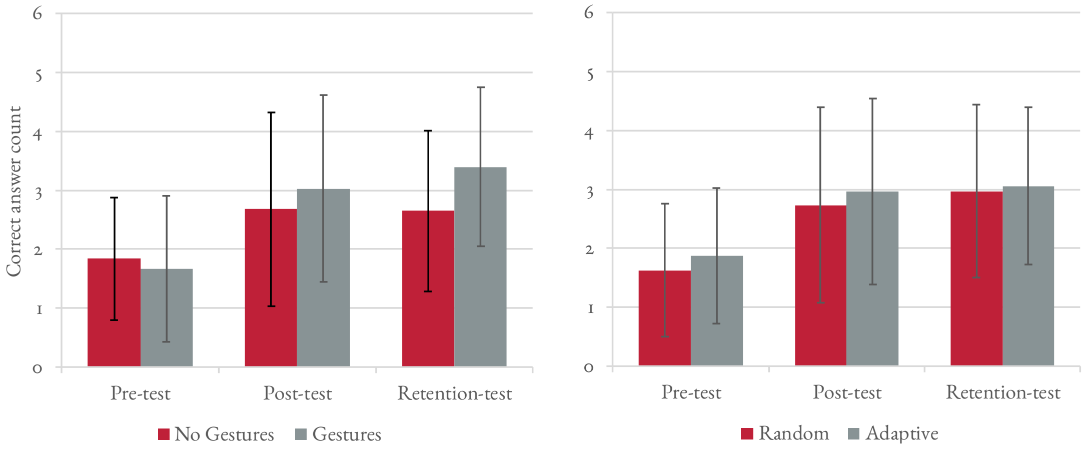


Figure 5.11: Test scores of the pre-, post- and retention-test. Left: Comparing gestures and no gestures condition; Right: Comparing adaptive and random condition (taken and redesigned with permissions from de Wit et al. (2018)).

also achieved a higher performance score ($M = .36$, $SD = .12$) as compared to those in the control (random) condition ($M = .32$, $SD = .06$), this effect of the tutoring strategy is not significant.

Consequently, the adaptive system alone was not able to improve the learning for young children, but when additional gestural support was provided, children in the adaptive condition turned out to perform even better ($M = .42$, $SD = .09$) than in the gestures only condition ($M = .34$, $SD = .06$). In fact, combining both yielded the best training performance ($F(1, 57) = 4.72$, $p = .03$, $\eta^2 = .06$). These results provide a first partial support for hypotheses H1 and H3 that children achieve greater learning gains when supported by iconic gestures and the adaptive A-BKT system during the training.

Subsequently, a second two-way ANOVA was used to analyze the previously calculated immediate learning gain between pre- and post-test (within-subjects factor) with respect to the use of gestures or tutoring strategy (between-subjects factor). Surprisingly, the results show no significant effects anymore (see Figure 5.11). Although the children in the iconic and adaptive conditions performed significantly better during training, this effect is not observable in their immediate learning gain. Furthermore, running the same ANOVA with the long-term learning gain as within-subjects factor also reveals no significant effect of tutoring strategy and, thus, hypothesis H4 is not supported. In contrast, for the use of gestures a significant effect was found ($F(1, 57) = 6.11$, $p = .02$, $\eta^2 = .097$), indicating a greater long-term learning gain between pre- and retention-test when gestures were present ($M = 1.70$, $SD = 1.56$) than without gestures ($M = .81$, $SD = 1.25$). However, this result provides just a partial support for hypothesis H2. Further, no interaction effect was found.

5.4.6 Discussion

In general, the evaluation results demonstrated that young children are able to learn new English words by spending just a single tutoring interaction of about twenty minutes with the developed SARTS. Additionally, they are able to retain this newly acquired knowledge for a prolonged period of time.

As expected, some children have been exposed to some of the target words beforehand, e.g., on television. This is reflected in the analysis of the pre-test, which indeed indicate a realistic amount of prior knowledge on average above chance. Nevertheless, post- and retention-test results are still higher and show the expected knowledge gain after engaging in the learning activities. Furthermore, it is noticeable that the post-test results are worse compared to the performance of the children at the end of the training stage. That is, instead of learning the concept of each word, the children might have just learned a simple link between a word being pronounced by the robot, a specific image and, in some cases, the iconic gestures produced by the robot. Although these results were not intended, they indicate that the changed test design with slightly different animal pictures allows for evaluating if the underlying concept of a word was learned. However, the tests still offer potential points of improvement. Although the children were explicitly instructed to select the image corresponding to the given English target word, sometimes they seemed to choose the animal with the most similar sounding Dutch name instead, e.g., "bird" was often confused with the Dutch word "paard". This might be resolved by integrating the target words into a meaningful context, e.g., by providing a semantic frame, instead of presenting them isolated (cf. Blasco, 2014).

Consequently, the results regarding the effects of the A-BKT model on young children's language learning are still inconclusive. One explanation for this might be the way in which the adaptive system teaches new words within the limited number of rounds, i.e., it focuses on the qualitative acquisition of words by providing extended support for those words that the child finds most difficult. However, the learning gain was measured as a quantification of newly learned words. Thus, learning with the adaptive system might not have resulted in a higher number of learned words, but in a deeper internalization of those taught words, which, however, was not measured in the tests. This might be resolvable by removing the upper round limit and ending the interaction at the point when the learning is assumed to be "optimal", i.e., the point at which the adaptive system believes the child has achieved the highest potential learning gain.

In addition, human teachers tend to personalize their learning content by drawing upon a memory that spans a longer period of time. In the current state of the experiment this is not feasible, since the memory of the adaptive system is just built up and applied over the course of one single session. But, having multiple sessions with each child and using the results of one session as prior knowledge for the next one could allow for the adaptive system to apply parts of teachers' personalization behavior and to unfold its full potential.

Another explanation might be that the decisions taken by the adaptive system, and, with that, the performed actions, were too subtle and, thus, remained unrecognized by the child, since only the order of words, their occurrence frequency and task difficulty was tailored to children's needs. Addressing also the perceived learning or allowing for more complex actions, for example, by introducing completely different tutoring strategies or making the difficulty levels more distinguishable, might enable the SARTS to fit the tutoring interaction better to a particular child.

Furthermore, it is possible that the A-BKT model has pushed the children into their ZPD, but could not provide sufficient scaffolding for them to succeed within this zone. This might have caused them to struggle during the learning process and prevent a deeper internalization of the vocabulary. This assumption is also supported by the study results, since children who learned with the adaptive system while being supported by additional iconic gesture achieved the best learning results during training. This, in turn, indicates that iconic gestures can be used to provide additional scaffolding for the cognitive learning dimension.

Finally, the post-test results might have been affected by a low engagement, which was observed for many children towards the end. Although this drop would also affect the performance during training, the influences might have been to small, since the performance score was calculated over the whole interaction. However, it is still possible that children would perform even better when playing with the adaptive system if they are highly engaged throughout the whole interaction.

In conclusion, the study demonstrated that children are able to learn new words from a SARTS applying the A-BKT model, but the chosen study design was probably still not optimal. Although it boosted the performance during training, the post- and retention-test results remained nearly unaffected, which might be due to children's low engagement at the end of the interaction. Thus, this is further investigated in the following.

5.5 Engagement Analysis

During study 2 a strong drop in the engagement for many children was observable, which seemed to be independent of the experimental condition. To verify and further analyze this subjective impression a perception online-study is conducted. Adult participants were asked to rate muted video clips showing tutoring interactions of study 2. The goal was to examine whether incorporating an adaptive tutoring strategy or iconic gestures can positively influence the change in engagement with respect to the following three hypothesis:

H1: Children's engagement will drop less when the robot uses iconic gestures compared to a robot not using the gestures.

H2: Children's engagement will drop less when the adaptive system is used compared to the random selection of skills.

H3: Children's engagement will drop least when the adaptive system and iconic gestures are jointly used by the robot.

5.5.1 Study Design

To compare the initial engagement level with the level at the end the participants had to watch video clips showing the fifth and twenty-fifth round of each child. These rounds are chosen, because they are close to the beginning and end of the tutoring interaction, but are not influenced by possible short bursts of engagement when children realize that the interaction with the robot is starting or ending. All clips are 5 seconds long and start when the robot asks the child to provide an answer to the given task. Within the scope of this perception study a previously excluded child with missing retention-test results is included, whereas another child is removed because of missing video data. This results in a dataset of 122 clips from 61 children, with 14 to 16 children in each condition. All participants have to watch the full set of clips in randomized order and rate the engagement level for each child on a 7-point scale, beginning from 1 (completely dis-engaged) to 7 (completely engaged).

5.5.2 Participants

A total of 21 participants took part in the perception study, from which three had to be excluded because of incomplete datasets. All remaining 18 participants were fluent speakers of the English language and had normal or corrected sight. Due to a break in the protocol, their gender was not and their age was only partially recorded. Analyzing the eleven datasets with existing age values revealed an average age of 26.14 years ($SD = 2.38$ years).

5.5.3 Procedure

To allow for the participant to practice the rating procedure two clips of children, who were excluded from the main test-set, are presented. One clip shows a situation in which a child was clearly engaged, whereas the other shows a clearly dis-engaged child. Subsequent to the practice round the participant receives a definition of engagement, i.e., responding rapidly to the questions, having an upright body posture, displaying joy after answering the question, and dis-engagement, i.e., responding slowly to the questions, supporting the head by leaning on the arms, showing less interest in the task, which are based on the cues visible in the video. After the participant has confirmed that she understood the task the actual rating procedure starts. After having assessed the engagement for all 122 clips, the study ends and the participant is thanked.

5.5.4 Results

To analyze the recorded ratings they have been averaged over all children of the same experimental condition. Since the participants rated videos of four conditions with two clips each, this results in a total of eight averaged ratings. The results are summarized in Figure 5.12. A paired-sample t-test shows that children are considered to be significantly more engaged in the fifth round ($M = 5.21$, $SD = .64$) compared to the twenty-fifth round ($M = 4.38$, $SD = .84$; $t(71) = -12.09, p < .001$).

To further examine the engagement course for the different study conditions, a two-way ANOVA was conducted based on gesture use (gestures versus no gestures) and tutoring strategy (adaptive versus random) as between-subjects factors and the relative drop ("round twenty-fifth minus round fifth") as within-subjects factor. The results reveal a significant effect of the tutoring strategy, $F(1, 68) = 86.26, p < .001, \eta^2 = .559$, so that the observed drop in children's engagement between the fifth and twenty-fifth round is smaller when playing with the adaptive tutoring strategy ($M = -.04$, $SD = .35$) as compared to a randomized word order ($M = -1.27$, $SD = .44$) and, thus, hypothesis H2 is fully supported. However, for the use of iconic gestures, as well as for the interaction effect between gestures and tutoring strategy, no significant differences are found, refuting hypotheses H1 and H3.

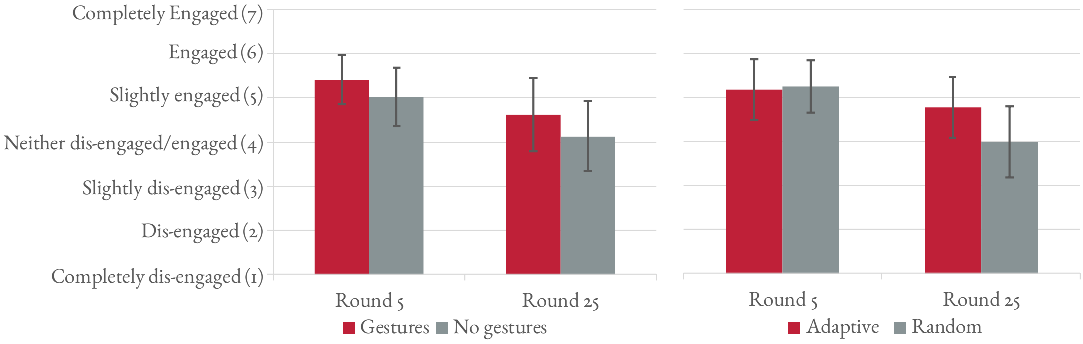


Figure 5.12: Engagement level ratings for the fifth and twenty-fifth round of the tutoring interaction. Left: Rating of the gestures versus no gestures condition. Right: Rating of the adaptive versus random condition (taken and redesigned with permission from de Wit et al. (2018)).

For a more detailed picture of the overall engagement in the different conditions throughout the entire training session the same analysis was repeated but with the average of the fifth and twenty-fifth rounds' engagement level for each condition as within-subjects factor. The results reveal that the overall engagement level is significantly higher in the gesture condition ($M = 5.02, SD = .63$) as compared to the condition without gestures ($M = 4.57, SD = .68; F(1, 68) = 8.75, p = .004, \eta^2 = .114$). Additionally, the engagement is significantly higher when using an adaptive tutoring strategy ($M = 4.97, SD = .67$) as opposed to a random strategy ($M = 4.63, SD = .67; F(1, 68) = 5.10, p = .03, \eta^2 = .07$), while again no interaction effect between both was found.

5.5.5 Discussion

In summary, the results of the perception study confirm the observed drop in engagement in all study conditions and, thus, highlight the importance of considering this aspect in the development of a SARTS. In general, learners' engagement is a crucial prerequisite for learning and can indicate a decreased motivation and willingness to learn (Blumenfeld et al., 2005). Incorporating additional information about the learner's affective state as part of the engagement or even the engagement itself into the A-BKT model can improve its impact not only on the engagement but also on the cognitive learning. In particular, if children are not in the right mood to learn, or their attention fades away, the SARTS could intervene to recreate or maintain the right atmosphere for learning and, with that, increase children's engagement level again (cf. Johansson, 2004). Although they might be able to succeed with lower engagement in a simple word learning interaction as presented in the previous studies, a high level of engagement might stimulate them to go beyond simple memorization and relate these new words to prior knowledge. Moreover, learners' engagement can serve as a measure of how well the learning activities are adapted to their abilities. For example, constantly presenting tasks that are either too hard or too easy can have a harmful effect on the engagement. First insights regarding this effect are observable in the presented study. While iconic gestures contribute to a higher overall engagement, the adaptive tutoring strategy resulted in a reduced decline in engagement towards the interaction's end. It appears that the adaptive system managed to tailor the learning activities to the needs of each child by providing a realistic, yet challenging task. Iconic gestures, instead, can be used as an additional scaffold for the learner to further support this process, since they achieved the best results during training when combined with the adaptive system (see Section 5.4.5). However, a constant use of gestures can lengthen the interaction and hamper children's engagement, which might explain the missing support for hypothesis H3. Instead of using iconic gestures all the time, it might be beneficial to turn them on adaptively, e.g., for more difficulty words, to provide an additional support while not lengthening the interaction unnecessarily.

5.6 Summary

This chapter focused on the question of how a SARTS can optimally address kindergarten children's cognitive learning (**RQ1**). This can, inter alia, be achieved by enabling the learner to work in the ZPD. However, this is not a trivial task and to achieve this, the SARTS needs modules to keep track of the learner's knowledge state (**RQ1.1**, *student model*), as well as a planning approach that enables the SARTS to adapt and individualize the learning interaction accordingly (**RQ1.2**, *pedagogical module*).

A promising approach to model the learner's knowledge is called BKT (Corbett and Anderson, 1994). To expand this approach with the ability to simulate and plan the next steps of the tutoring interaction, an additional action decision node was added resulting in a DBDN-like model called A-BKT. It allows, for instance, to focus on a smaller set of skills, with which the learner is struggling before overwhelming her with too much new content. As a first step, the included actions just describe abstract task difficulties, which are mapped on concrete exercises in the SARTS architecture at a later stage. Consequently, the A-BKT model combines the *student model* and the *pedagogical module* of an ITS and, additionally, is easily extendable to incorporate further aspects of the remaining learning dimensions later on.

The proposed model was evaluated in two user studies, first, with adults to check the rudimentary functionality without having to cope with additional difficulties arising from a cHRI, and later with kindergarten children. Although the results of both studies showed first benefits when applying the developed adaptive tutoring model, they are still inconclusive. While a positive effect of the A-BKT model was found during the training stage of the interaction, especially when paired with iconic gestures produced by the robot, this effect was nearly absent in the post-tests.

Different explanations might account for this finding. Apart from possible deficits in the study design the tasks might have been too easy or too difficult for the learners, so that they either did not work in the ZPD or could have needed further scaffolding. The latter might also explain why a strong interaction effect between adaptive system and iconic gestures was found during the training stage of the game. There, the iconic gestures might have worked as an additional scaffold for the learner to succeed in the provided learning tasks.

Thus, providing additional scaffolding in all dimensions of learning and the engagement might further support the effect of the adaptive model, which in turn could also result in more differentiated post-test results between the conditions. In fact, especially in study 2 a strong decline of engagement towards the end was observed. Since the learning dimensions are strongly interconnected, providing an additional scaffold to maintain a constantly high level of engagement as well as addressing the affective learning could influence the cognitive learning and, consequently, also the learning outcomes (see Section 2.1.6). In addition to the possibility to directly address the engagement and affective learning, it can further be improved by the perceived learning. That is, if the learner attains a good self-perception this can further foster her learning motivation and, hence, her concentration and engagement.

In conclusion, the developed SARTS based on the A-BKT model already provided promising results during foreign language learning sessions, but requires further refinement and extensions to yield its full potential. One option could be to introduce further scaffolding actions or strategies, which are applicable by a SAR to support the learner in each learning dimension. Hence, the following chapter presents work on different scaffolding approaches for the affective and perceived learning dimensions, as well as the learner's engagement.

What a child can do today with assistance, she will be able to do herself tomorrow.

— Lev Vygotsky

6 Scaffolding Affective and Perceived Learning, and Engagement

In the previous chapter an adaptive approach called A-BKT was developed that allows to address the learner's cognitive learning by tailoring the tutoring content to their individual needs. The evaluations with young children demonstrated that the A-BKT model can support their learning gain and engagement, but they also provide hints that further scaffolding is required. Scaffolding is not limited to the learner's cognitive learning and can also address the remaining learning dimensions, as well as engagement (**RQ2**). Consequently, this chapter investigates what are relevant behavioral cues to track the engagement of kindergarten children (**RQ2.1**) and which actions can be used by a SARTS to scaffold the engagement and affective learning during foreign language learning interactions (**RQ2.2**). Moreover, since scaffolding can also support learners' perceived learning dimension, this chapter further examines which strategies could be applied to scaffold young children's perceived learning during the learning interaction (**RQ2.3**).

To address these questions this chapter first describes how a SARTS can be enabled to track and manipulate learners' engagement and affective learning, so that it can provide appropriate scaffolding for them (Section 6.1 & 6.2). Second, a scaffolding technique to support also children's perceived learning is derived and implemented. Subsequently, the findings are integrated into the SARTS and evaluated with children in German kindergartens (Section 6.3).

6.1 Engagement and the Dimension of Affective Learning

As already mentioned, the learner's engagement plays a crucial role in learning interactions. Together with affective learning both strongly influence the cognitive learning. Motivational prior beliefs, as well as affective states, such as boredom, flow and confusion have been shown to either positively or negatively influence learning outcomes (Bandura, 1991; Craig et al., 2004, see Section 2.1). Consequently, providing appropriate scaffolding by motivating the learner to continue, preventing boredom and resolving confusion can result in a better learning flow, as well as in a more beneficial tutoring interaction for the child. While this type of scaffolding does not influence the learning progress directly, it still supports the learner in succeeding within the learning setting, e.g., by encouraging to go on and to try again after failing a task.

However, within this thesis not all levels of the affective learning dimensions needs to be addressed. That is, the learner should perceive, think about and respond to the learning content, as well as develop a positive value towards the learning interaction in general. In contrast, the ability to organize the learning time and to establish an own learning profile are not relevant within this thesis, since kindergarten children still lack the required cognitive development (cf. Peace Corps, 1986, p. 67f.) and, thus, they have been excluded for now (cf. Krathwohl's taxonomy, Section 2.1.2).

One option to address these aspects of learning, at least partially, is to provide appropriate feedback with regard to the learner's performance during the learning interaction. Therefore, an empirical basis was collected and analyzed, inter alia, with respect to educators feedback behavior (see Chapter 4). It serves as a basis to derive and implement feedback strategies for the SARTS, which are in line with daily pedagogical practice.

However, the use of appropriate feedback, which mainly focuses on the learner's intrinsic motivation, is not the only way to support their affective learning domain. As mentioned in Chapter 2, also learners' deeper cognitive processes and affective states as included in the engagement are important, whereas each process and state probably require different actions each. This makes a classification of the learner's behavior indispensable to allow for a SARTS to autonomously provide appropriate support. But since internal states, e.g., thinking more deeply about the learning content (cognitive engagement and affective learning), usually can not be tracked directly in a way applicable in real life settings, tracking systems often rely on the shown behavioral engagement and affective states (see Section 2.1.4). This is probably also the reason why the reactions or scaffolding actions of ITSs are often designed to mainly address just these particular aspects, since it allows for the system to observe the effects of actions and, thus, to adapt the system's behavior correspondingly. However, since all parts of learning are strongly correlated, the mentioned strategies can also influence the hardly traceable parts of the affective learning and engagement.

In conclusion, besides of applying appropriate feedback strategies, the SARTS also needs to track and manage the learner's behavioral engagement, whereas the learner's affective learning and cognitive engagement is assumed to be further affected through the implicit influences of the executed actions.

6.1.1 Engagement Detection

In general, two different directions to track the user's engagement automatically can be found in the literature of Human-Machine Interaction (HMI). First, directly tracking the engagement, e.g., by training classification systems on huge datasets or, second, focus on tracking the affective and cognitive states that are known to be indicative of the engagement and affective learning (see Section 2.1.4 & 2.1.2).

These states can then be used to make decisions for the course of the interaction (e.g., D'Mello and Graesser, 2012) or to infer the engagement later on (e.g., Altuwairqi et al., 2018).

Most of the commonly used approaches rely on machine learning techniques to automatically analyze children's behavioral cues and to derive their corresponding affective and cognitive states, or engagement level. To capture the different modalities, such as facial expressions, body postures, the voice or information from the interaction itself, sensory inputs (e.g., from cameras and microphones) are used. But often these techniques, e.g., Support Vector Machines (SVMs), are supervised and need preprocessed input data, where the input values are transformed into numerical representations before each datapoint is labeled, e.g., with the corresponding engagement level. With this, a SVM, for instance, can learn the mathematical functions to map the input data to the corresponding output labels, which then can be used to classify the user's engagement level from new input data during an interaction. Although this process increases the time needed to prepare the machine learning environments and can increase the difficulty to transfer the approaches to different contexts, it still is often used in the community.

Castellano et al. (2012), for instance, used this technique to directly assess the user's engagement based on pre-recorded sessions of a chess game interaction. They compared several SVM-based models trained with different information about the game (e.g., game state and game evolution), the social context of the interaction and the game turn level (e.g., encouraging comments and scaffolding). Their evaluation showed that the SVM incorporating the social game context, as well as interactional information about the game, achieved the best performance with an average accuracy of 80%. However, basing the classification process on specific game state information can complicate its generalization and transfer to other games/contexts.

A different approach from Sanghvi et al. (2011) used the postural expressions of children to train five other supervised classification approaches in Weka[1] (ADTree, OneR, LogitBoost, MultiClassClassifier and logistic functions). The postural expressions data consist of features of users' full upper body silhouette that were automatically extracted by a computer vision algorithm and, subsequently, labeled by three trained coders. Their results demonstrated an accuracy of up to 82% for two of the five classifiers, namely, ADTree and OneR.

Also many other approaches can be found to directly classify users' engagement, which incorporate conditional random fields, e.g., by using audio-visual data (Foster et al., 2017) or personalized deep learning frameworks, e.g., by using audio-visual and physiological data (Rudovic et al., 2018). However, the majority of approaches seem to focus on only tracking users' affective or cognitive states.

Since the cognitive states mainly describe the internal states that are often not expressed through overt behaviors, most research focuses on tracking just the user's attention. This is generally done by analyzing the user's gaze behavior during the interaction, e.g., directly with an eye-tracker (El Haddioui and Khaldi, 2012; Yang et al., 2013) or the head pose (Lemaignan et al., 2016a). For example, Lemaignan et al. (2016a) used the OpenCV[2] framework to calculate children's attention and "with-me-ness" during

[1] https://www.cs.waikato.ac.nz/ ml/weka/index.html

[2] https://opencv.org/

a robot-teaching task. Here, the term with-me-ness represents a measure of "how much the user is with the robot during a task" (Lemaignan et al., 2016a, p. 163). Their evaluation demonstrated for both aspect results fairly close to the used ground truth provided by human raters.

In the scope of affect detection, however, a broader variety of different modalities and approaches were already used. One widely applied method, also for commercial products, such as Affectiva Affdex (McDuff et al., 2016), is the analysis of facial expressions to detect the affective state of a user (e.g., McDaniel et al., 2007; Wang et al., 2018). But these classifiers are often trained on "very expressive and acted" emotions for which they yield classification accuracy around 97%, however, this makes their applicability to real-world interactions questionable (cf. Stöckli et al., 2018). In fact, the accuracy of emotion detection based on facial features is often low in real-world applications and the recognition rate is strongly dependent on the individual expressiveness of each person (cf. Benta and Veida, 2015; Stöckli et al., 2018).

An alternative approach is the detection of affective states from the user's voice (Devillers and Vidrascu, 2006; Kim et al., 2017; Tzirakis et al., 2018). Classifiers to analyze the voice are often trained on datasets of spontaneous speech, so that they are more suitable for real-world applications. Kim et al. (2017), for instance, used a SVM with the aim of assessing the affective states of users interacting and playing with robots, such as Robotis OP2, Robotis Mini or Romo. To train their model they used the IEMOCAP database[3], which contains approximately 12 hours of audio-visual data from adult speakers and their evaluation showed a reasonably good classification performance (Kim et al., 2017). However, with regard to a cHRI setting, affect detection through speech analysis is difficult, because speech input is not always included since ASR systems for young children still have low accuracy (Kennedy et al., 2017b).

Other attempts made use of analyzing written text to detect the user's affective states (Alm et al., 2005; Kahn et al., 2007), which include, for instance, analyzing the usage of adjectives and adverbs. However, this approach is not applicable for kindergarten children, since they are usually not able to read and write text, while even adults usually do not use written text in natural face-to-face interactions.

A more extensive approach for affective state detection is the tracking of the whole body posture and movements, e.g., by using a Microsoft Kinect (McColl and Nejat, 2012) or a body pressure mat laying on a chair (D'Mello and Graesser, 2009). A limitation of the latter is that it assumes the user stays seated without moving around, which is also not easy for young children in the age of 4-6 years. The Kinect, however, allows the user to move around, but may have problems in detecting smaller events, such as small gestures or postural shifts.

In addition, approaches based on human physiology have been developed, too. In this realm, techniques, such as ECG, EEG, EMG (Wagner et al., 2005; Villon and Lisetti, 2006), and brain imaging (Immordino-Yang and Damasio, 2007) are applied to "read" the affective state from users' physiological signals. The results of these methods are promising, however, the applicability of such obtrusive approaches (e.g., wires and patches on the body) in tutoring interactions with young children is clearly limited, while the interaction itself is also not very natural anymore.

[3] https://sail.usc.edu/iemocap/

Finally, also multi-modal approaches were studied with the goal to overcome some of the previously mentioned limitations and to increase the detection accuracy. A lot of combinations can be found that incorporate, e.g., facial expressions and voice (Busso et al., 2004; Esposito, 2009), facial expression, voice and body posture (Bänziger et al., 2009), facial expressions, body postures and context dependent activity logs (Kapoor and Picard, 2005), or speech and text (Arroyo et al., 2009). And indeed, most of these systems demonstrated that a multi-modal approach to detect affective states results in higher accuracy rates.

In summary, a lot of different approaches with a broad variety of used modalities have been developed so far, which, however, are not all applicable when trying to provide an interaction as natural as possible for young kindergarten children learning with a SAR. Most of them have special requirements, are very obtrusive or have a low accuracy when applied in the wild. Furthermore, they often have not been verified to work with young kindergarten children and, consequently, are not necessarily applicable for this particular age group. In fact, observations of study 2 revealed that most children do not provide much speech input nor show a lot facial expressions during the interaction with a SARTS. Instead, approaches that are based on learners' gaze direction, body posture or movement might be successful in tracking their engagement. However, most affect detectors are trained on a huge amount of specifically annotated data to identify the important cues for each affective state, which also holds for commercial products such as Affectiva Affdex. Establishing such a dataset in a natural way without using an artificial setting for data recording is very complicated and time consuming, especially when not having a clue which features are important. Moreover, due to the fast development of children (cf. Mooney, 2013, Chap. 4), a dataset for different age ranges might be needed to train separate classifiers to be able to autonomously and reliably track children's engagement based on their behavioral cues.

6.1.2 Affective Tutoring Systems

Although the engagement detection for young children still yields complicated challenges, a broad variety of approaches demonstrated the general feasibility with reasonable accuracy. Because of this technical progress, so-called Affective Tutoring Systems (ATSs) aim for including this information into the *student model*, which, in fact, can increase student's learning performance up to 91% compared to a non-affect-aware ITS (Shen et al., 2009).

Alexander et al. (2006), for instance, developed an ATS called Easy as Eve including a virtual agent for primary school students. The affective states are detected by analyzing the facial expressions of the student and serve as basis for the selection of the next tutoring actions. To achieve this, case-based rules were defined and applied that have been informed by an observational study of human tutors beforehand. In an evaluation study conducted in a primary school, in which children had to solve mathematical equations, the use of their affective system showed a significant increase of students' performance as compared to a control group without affective support (Alexander et al., 2006).

Another system called the Affective AutoTutor can automatically detect boredom, confusion, frustration and neutral affect by monitoring conversational cues and discourse features along with gross body language and facial features. The cues provided by each "channel" are combined to select a single affective state, based on which AutoTutor responds with empathic, motivational, or encouraging dialog moves and emotional displays. The evaluations showed that students, whose learning is supported by this system, are not only able to pick up new knowledge but also to apply it in transfer tasks later on (D'Mello and Graesser, 2012).

Gordon et al. (2016), instead, incorporated a valence and engagement detection via facial expressions based on the Affectiva Affdex framework for their cHRI. In a study with preschool children they showed that their system personalized its policy over the course of training and that children who interacted with the personalized robot showed higher long-term positive valence as compared to a control group without personalization (Gordon et al., 2016).

In summary, promising findings from a variety of approaches support the inclusion of engagement or at least affect detection in ITSs. In addition to the ATSs mentioned above, many more can be found that try to cope with learners' affective states or engagement, respectively, e.g., the Cognitive Tutor Algebra (d Baker et al., 2012), Wayang Tutor (Wixon et al., 2014), VALERIE (Paleari et al., 2005) or Guro Tutor (Olney et al., 2012). However, developing and training an appropriate classifier applicable for kindergarten children is still a complicated task, especially with the goal of keeping the tutoring interaction as natural as possible. But even with a fairly accurate engagement classifier, suitable actions are still needed to respond to the respective states. They can be used to provide appropriate scaffolding for the learner to prevent the loss of her learning motivation or to re-engage her if necessary. However, since these actions and strategies aim for manipulating children's emotions and/or internal states, they have to be designed carefully. One option is to base their development on data of observational studies of human tutors and their reactions to behaviors shown by their students (cf. Alexander et al., 2008), which further allows to narrow down the feature space for the training of a classifier. Since to the best of the author's knowledge such a dataset does not exist for the respective age group of 4-6 years, an own empirical basis has to established. It should allow to inform the SARTS so that it is enabled to detect changes in children's engagement or the respective learning-relevant affective and cognitive states, respectively, and to react to these changes appropriately. Therefore, it should provide suggestions for possible actions to be performed by a SAR, or a SARTS in general, either pro-actively or as repair mechanisms when the engagement and learning motivation is already lost. Furthermore, it should be usable as a basis to inform a Wizard of Oz (WOz) (Dahlbäck et al., 1993) or to develop an engagement tracker later on.

6.2 Study 3: Empirical Basis for Detecting and Managing Engagement

To establish an empirical basis that includes children's behavioral cues to track the engagement, as well as appropriate actions to manage it, specific knowledge from experts (educators) in reading and managing the affective and cognitive states of kindergarten children in tutoring interactions is required. To collect their knowledge about which affective and cognitive states occur and are important during a tutoring interaction, and how they can be detected just by observing the child while playing with a SARTS a qualitative approach is chosen. Video recordings of cHRIs from study 2 are used to interview

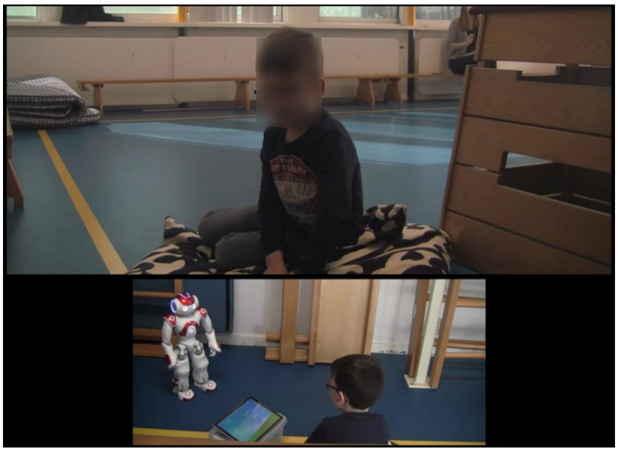


Figure 6.1: Anonymized screenshot from one video shown to the experts in the interviews displaying a learning interaction from two perspectives (reprinted with permission from Schodde et al. (2017b)).

educators on their perception and interpretation of children's behavior. Subsequently, the recorded interview sessions are transcribed to enable a detailed content analyses of experts' comments with respect to the following research questions:

RQ2.1: What are relevant affective and cognitive states, as well as behavioral cues, to track the engagement of kindergarten children during foreign language learning interactions?

- **RQ2.1a:** How do experts interpret the affective and cognitive state of children during the robot-child tutoring lessons?
- **RQ2.1b:** To which behavioral cues do they refer when they remark changes, e.g., in the child's level of attention?

RQ2.2: How would the experts react to changes in children's engagement from the perspective of the robot?

6.2.1 Study Design

Video recordings of study 2 with Dutch kindergarten children are used for the interviews to enable the experts to observe children's behavior during an interaction with a SARTS in a controlled manner. To ensure that individual differences are also considered despite of a small sample size, video recordings of eight different children (4 female, 4 male, 1 video per child) are taken, who varied in their level of activity and expressiveness when learning with the SARTS. Furthermore, these videos are shortened in time by removing some rounds in which children showed the same behavioral cues or no cues at all. This allows for the experts to judge as many children as possible (four), while not overstraining them with regard to their time and concentration. To enable them to get a good impression of the children's behaviors, the interaction is presented from two different camera perspectives, a frontal view of the child, as well as an overview from the side showing the whole experimental setup including the robot, the tablet and the child (see Figure 6.1). The final set of videos has an average length of 11:18 minutes (*SD* =22 seconds) and is presented and discussed during face-to-face interviews.

Number	Gender	Age	Experience
1	female	45 years	27 years
2	female	51 years	35 years
3	female	50 years	25 years
4	female	36 years	16 years
5	female	61 years	42 years

Table 6.1: Profile data of all kindergarten teachers, who participated in the interviews as experts.

To record the whole interview session a computer microphone and a screen capture tool are used. This allows for synchronizing experts' comments with the video recording that is shown on the screen at each point in time.

6.2.2 Participants

A total of five kindergarten teachers were invited and interviewed as experts, who cover a broad range in age and working experience (see Table 6.1). They were between 36 - 61 years old ($M = 48.60$; $SD = 8.16$) and had a working experience ranging from 16 to 42 years ($M = 29.00$; $SD = 8.88$) . They are all native German speakers and had normal or corrected sight.

6.2.3 Procedure

At the beginning of each interview session the expert is informed about the purpose and procedure of the interview and had to sign an informed consent that her voice will be recorded. She is instructed to judge children's behavior and related affective or cognitive state, which are presented in the video recordings on the laptop. First, the expert sees a short example video, which she has to comment on to make sure the task is clear. Subsequently, the first video of the main commenting process starts, which is guided by a few example questions, requesting the expert to estimate whether the child is attentive, bored and/or has fun while playing with the SAR. If not explained directly, the interviewer asks the expert to justify her decision to get further information about possible behavioral cues.

Subsequent to each video the interviewer asks how the expert would react to the observed negative changes in the child's state. For example, what would she do if she recognizes a lack of attention and how the resulting actions could be realized with a SAR. At each point in time, the interviewee is allowed to pause the video and go back to review a scene. Overall, each expert discusses a total of four videos with the interviewer before she is thanked for her participation and dismissed.

6.2.4 Analyses and Results

Analyzing the experts' descriptions and explanations revealed different categories of affective and cognitive states (**RQ2.1a**) with the corresponding behavioral cues for their identification (**RQ2.1b**). As depicted in Table 6.2, the listed states can be grouped into the meta-level states of engagement, dis-

Meta-level State	**State Interpretation**	**Behavioral Cue**	**n***
Engagement	Concentration/ Thinking	eye contact	5 (4)
		sit still	2 (2)
		hand to head	4 (3)
	Involvement/ Activity	mimic robot's gestures	2 (2)
		answer verbally	1 (1)
		nodding	1 (1)
		head-shaking	1 (1)
	Expressive/ Proud	smiling	7 (4)
		thumb up	1 (1)
		raise fist	1 (1)
Dis-engagement	Inattentiveness/ Distraction	rub eyes	2 (1)
		grimace	4 (4)
		gaze away	7 (4)
		turn away (whole body)	10 (4)
		change position (stand up & lay down)	2 (2)
	Boredom/ Impatience	support the head with hand(s)	3 (2)
		move the head from left to right	2 (2)
		undirected finger tapping	4 (3)
		gaze away	2 (1)
		move position (stand up, lay down)	6 (4)
Negative Engagement	Skepticism	tilt head	3 (3)
	Disinterest	frown	1 (1)
	Averseness	lower mouth corners	1 (1)

*n is the frequency of reference to a cue; the number of children for which the cue was observed is noted in parentheses.

Table 6.2: Behavioral cues shown by children during the interaction, their related state interpretations and the corresponding meta-level states (reprinted with permission from Schodde et al. (2017b)).

engagement, and negative engagement. The engagement comprises states, such as concentration and thinking, activity and involvement, as well as expressiveness. That is, if children keep eye contact with the robot and/or tablet and sit still, the experts interpreted their behavior as concentrated and engaged. Further, if they reenact the robot's iconic gestures, or answer verbally or nonverbally, e.g., with nodding or head-shaking, they are also described as involved and engaged in the interaction. Likewise, expressive behaviors, such as smiling or showing a thumb up, are also interpreted as signs of engagement by the experts. In contrast, behaviors that are interpreted as signs of distractions, inattentiveness or boredom are regarded as indicators for dis-engagement. For instance, if children were rubbing their eyes, gazing away from the setting or changed their seating position frequently, they are classified as inattentive. Additionally, supporting one's head with the hand(s), showing undirected finger tapping, gazing away from the robot and/or tablet, etc. are named as noteworthy behaviors that indicate boredom and dis-engagement (cf. Table 6.2). Finally, the last category called negative engagement is composed of negative states, such as skepticism and averseness towards the tutoring interaction with the robot. These states are related to behavioral cues, such as frowning, lowering the mouth corners and head-tilt.

Preventive actions	**Paraphrases**	**n***
Include verbal input	It would be more motivating for the child if it should talk to the robot (expert 2, video 2)	3
Heighten robot's activity (e.g., move head)	The interaction would be more engaging if the robot moves. (expert 2, video 2)	3
Repair actions		
React to the child's behavior/ give feedback	The robot should react to the behavior of the child, e.g., tell him/her to sit down again. (expert 5, video 1)	4
Change task difficulty	The task should increase in difficulty to get the child's attention back. (expert 1, video 3)	1
Include alternative activities (e.g., play a game; stand up)	The robot could ask the child to stand up and move around, so that he/she is ready to listen again afterwards. (expert 3, video 2)	4
Allow a break	A break or a continuation at another day could be helpful to get the attention back (expert 2, video 1)	2

*n is the number of experts out of the 5 experts that mentioned the strategy.

Table 6.3: Possible (preventive) actions to address children's engagement mentioned by the educators (reprinted with permission from Schodde et al. (2017b)).

Since each interaction with the robot varied with respect to the individual differences of the children (e.g., in age and self-confidence), the frequency of how many times each behavioral cue was mentioned by different experts for different children is counted. For instance, when two experts judged a cue for one child as relevant, it is counted as two, whereas when a cue was mentioned by only one expert but several times for one child, this is only counted as one occurrence (see Table 6.2). Moreover, to allow for a reflection on the occurrences of a particular cue over different children, further the number of children for whom this cue was observed is counted (see Table 6.2 numbers in parentheses).

The final counts indicate that behavioral cues, such as eye contact ($n = 4$ children), smiling ($n = 4$), and self-touches to the head ($n = 3$) are interpreted as a sign of engagement for multiple children during the interview. Regarding dis-engagement, cues, such as making grimaces ($n = 4$), gazing away ($n = 7$) or even turning away from the robot and/or tablet ($n = 4$), moving the position ($n = 2$) and undirected finger tapping ($n = 3$) are observed across several children. As a sign of negative engagement head tilt is named as a behavioral cue used by several children ($n = 3$) to express skepticism. In contrast, cues, such as giving verbal answers, nodding, head-shake, eye rub, frowning, and lowered mouth corners, are only observed for one child and, hence, appear less informative.

In addition, the experts were asked how they would intervene to keep children engaged in the interaction from the robot's point of view (**RQ2.2**). Their suggestions are summarized into categories of potential actions to re-engage children who learn with a SARTS (see Table 6.3). Some of the experts' suggestions can be regarded as preventive strategies that can be employed in the interaction from the outset (preventive actions, see Table 6.3). These are general strategies, such as allowing multi-modal interactions (e.g., requesting the child to provide speech input) or more expressive and varying robot behaviors (e.g., gestures and movements) to keep children engaged in an interaction. Beyond that, the experts also mentioned actions, which can be useful to re-engage children in an ongoing interaction after an engagement drop is observed (repair actions, see Table 6.3). For example, if only first signs of dis-engagement are recognizable, the robot can, e.g., increase the task difficulty. Instead, in cases of higher dis-engagement, the robot can suggest alternative activities to get the child's attention back, e.g., by playing another game on the tablet. However, according to the experts' opinions, in some cases it will even be necessary to pause the tutoring completely and to provide a break for the child, e.g., by standing up and do some physical exercises.

In summary, this empirical basis can serve as a basis to track the learner's engagement and to develop and implement appropriate re-engagement actions executable by a SARTS. For the former aspect, the identified behavioral cues, and affective and cognitive states can be used either to train a WOz or to narrow down the feature space for building a dataset to train an engagement classifier in the future. Therefore, within this thesis the term of engagement is defined as follows:

> Engagement refers to the collaboration between child and SAR and relates to the learner's affective and cognitive states, namely, concentration, involvement, joy, attention and interest towards the tablet game and the robot's behavior. Hence, dis-engagement is defined as inattentiveness, boredom, frustration and impatience.

6.3 Study 4: Effects of Explanation and Re-Engagement Strategies

Within the scope of the following study two different scaffolding strategies are evaluated with young kindergarten children. First, to address the children's perceived learning it is examined whether kindergarten children can already benefit from a SAR that reveals its current beliefs about the child's knowledge state and, based on that, explains the resulting changes of the language learning interaction. This transparency, especially with respect to the learner's already attained knowledge, is intended to support and guide the process of self-perception, as well as the attribution of errors. This can be interpreted as a preparatory step towards the acquisition of basic abilities required for SRL and can result in better learning outcomes. Further, it can be expected that these explanations will make the system's adaptations more salient and, with that, foster not only engagement but also perceived and cognitive learning (cf. Section 5.4). Finally, it can be assumed that this strategy leads to a perception of the robot as a more competent and reasonably acting companion, which can have a positive effect on the acceptance of the whole learning environment (see Section 3.3.2).

Second, this study aims for investigating whether a child can be kept engaged or re-engaged during the tutoring interaction with a SARTS. To achieve this, the identified preventive strategies for maintaining engagement, as well as the possible re-engagement actions are included into the system (see Section 6.2.4).

Summarized, the study aims for investigating the following hypotheses:

H1: Children who interact with a robot tutor that explains its belief about their knowledge are more motivated to continue learning than children who interact with a robot that does not provide explanations.

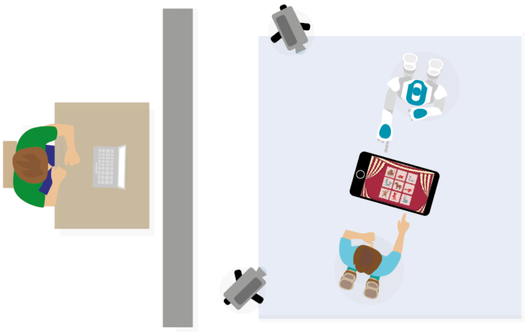


Figure 6.2: Experimental setup of the robot–child tutoring interaction supervised by technician (reprinted with permission from Schodde et al. (2019)).

H2: Children who learn with a robot tutor that explains its belief about their knowledge show a higher learning gain than children who interact with a robot that does not provide explanations.

H3: Children who learn with a robot tutor that explains its belief about their knowledge gather a better understanding of their perceived learning than children who interact with a robot that does not provide explanations.

H4a: Dis-engaged children will be re-engaged through dedicated actions executed by the robot.

H4b: Dis-engaged children will be more re-engaged when the robot explains its decision to start a re-engagement attempt.

6.3.1 Study Design

To study these aspects a between-subjects design is employed in which the children either interacted with a robot that explains its decisions and beliefs about the children's knowledge before the target word is announced or not. However, the engagement related actions are used in both conditions equally.

The general study setup is again based on the empirical basis described Chapter 4. It is supervised by two experimenters, who stay out of the children's field of view, whereas one controls the robot and the other observes the children. Moreover, the whole interaction is recorded from different angles by two video cameras to be able to analyze the children's behavior later on (see Figure 6.2).

Although the basic setup remains the same, some further adaptations are made to improve the interaction with respect to educators' suggestions and findings of previous studies.

6.3.1.1 Scenario and chosen vocabulary

While the underlying interaction structure of playing "I spy with my little eye ..." remains the same, the setting in which this game is played has changed to also include a small background story. The new setting is based on a virtual circus visited by the children together with the robot. While a circus is commonly known and liked by children (cf. Jampert et al., 2006), it also provides a reasonable frame for different animals showing up at the same place.

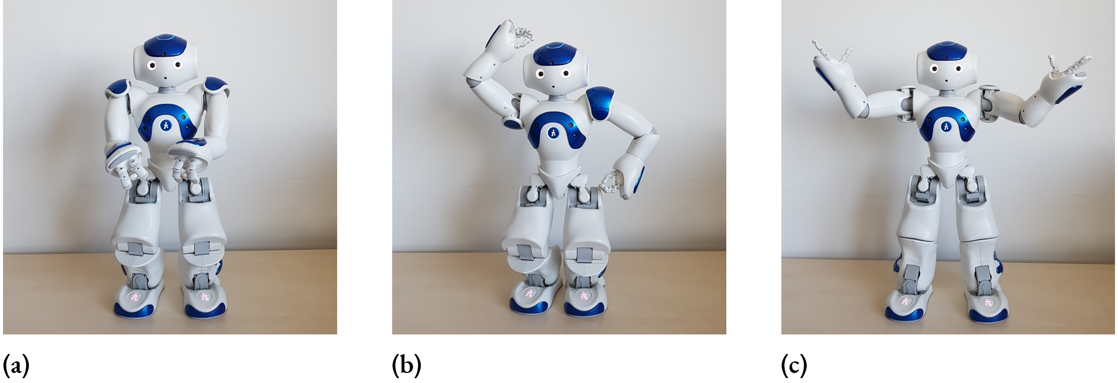


Figure 6.3: Examples of iconic gestures produced by the robot for the animals used in the experiment. (a) Imitating the riding of a horse by grasping imaginary reins and moving the arms up and down; (b) imitating a monkey by scratching armpit and head; (c) imitating a lobster by raising the arms while opening and closing the hand like lobster-claws (reprinted with permission from Schodde et al. (2019)).

Furthermore, to challenge but not to overwhelm the children with the amount of vocabulary, a collection of nine English animal names is used as target words. In comparison to study 2, the number of target words is raised again, because the A-BKT system revealed only a small positive effect on the children's learning gain. An assumption is that the difference between the three difficulty levels (just 1, 2 or 3 distractors) was too small. With more target words it is possible to create more salient differences (e.g., 2, 5 or 8 distractors), which allows for a better fitting to the proficiency level of each child. Moreover, a high number of target words enables the children to learn more words, which is especially important for children with prior knowledge. For example, if a child already knows 3 out of 6 target words, just 3 further words can be learned. However, with 9 target words this number is increased to 6 unknown words.

In addition, the set of used target words has been adapted, too. All included animals are carefully chosen with regard to theoretical assumptions and lessons learned from earlier experiments (cf. Section 5.4). For example, some animal names, such as "chicken", were found to be commonly known and, thus, had to be exchanged. Further decisive factors taken into account are that (1) the animals should offer the possibility to create a comprehensible iconic gesture executable by the robot, e.g., riding on a horse (see Figure 6.3a), (2) the names should be as different as possible between the German and English language, and (3) it should be possible to create meaningful groups of animals with the same color. The final set of target words for the tutoring interaction comprises a parrot, a ladybug and a lobster as red-colored, a horse, a bull and a monkey as brown-colored, a rabbit, a seal and a snake as gray-colored animals (see also Figure 6.5d). Further, to ensure that all animal names are pronounced correctly by the robot's Text-To-Speech (TTS) engine, they are spelled in the phonetic alphabet.

Moreover, since the differences between the previously used task difficulty levels were assumed to be too subtle and, thus, remained unrecognized (cf. Section 5.4.6), they are reworked with respect to the new number of target words so that they are more distinguishable. Further, an additional level of difficulty is added, which considers the animals' colors as an additional influence factor. Especially with the underlying mechanics of the used tutoring game, colors are very important since they are a salient feature of objects (cf. Barrow et al., 2000) and, thus, might result in children remembering the animal's color instead of the target word. Based on these changes the following four levels of difficulty from easy to hard can be created: 3 animals with 3 different colors, 3 animals with the same color, 6 different animals (2 red, 2 gray, 2 brown), and 9 different animals.

6.3.1.2 Actions to prevent dis-engagement

To be in line with experts' suggestions regarding actions to prevent dis-engagement (see Section 6.2.4), small rewarding feedback behaviors executed by the robot are implemented, e.g., gazing towards the child or nodding when verbal input is received. To underline the positive feedback given by the robot, each time the child answers correctly the robot's eyes are blinking in rainbow colors to resemble a smiling face (cf. Fridin, 2014b). These behaviors are intended to let the robot appear more "alive" and the interaction more natural, which in turn can slow down the loss of children's engagement.

Furthermore, iconic gestures are used to provide multimodal input for the child, which already demonstrated their positive effect on children's engagement in study 2 (see Section 5.4). However, this time the gestures are executed by the robot only once for each animal during their introduction and not each time when an animal name is mentioned during the training stage of the game. This should prevent an unnecessary prolongation of the tutoring interaction, which might negatively influence children's engagement and with that the study's measurements.

Finally, the underlying adaptive system (A-BKT model) is allowed to terminate the interaction early if a child already learned most or even all of the target words (average belief state $>= 75\%$). This should prevent children from repeating already known words over and over again, which could negatively impact their engagement and interfere with their learning gains.

6.3.1.3 Repair actions to recover engagement

In addition to the preventive actions executed by the robot regardless of the learner's state, additional repair actions are designed. These can be carried out spontaneously during the interaction if engagement drops are observed, since, according to educators, the robot might be able to get back the child's attention by noticing, acknowledging and reacting to a decrease in engagement (see Section 6.2.4).

But developing a sophisticated classifier for engagement detection is a complex and time consuming task that often requires a huge labeled dataset for the training process. Further, a low classification accuracy might cause misclassifications, which in turn can result in inappropriate robot behavior. To prevent this and to provide a tutoring interaction as consistent as possible, a human WOz is trained, who keeps track of the children's behavior. To this end, the most informative cues derived from the conducted expert interviews to detect dis-engagement serve as a basis (see Table 6.4). These cues were sorted into meaningful groups, namely, tired, heightened activity, low and high distraction, which are selected by the WOz as soon as the child shows one of the respective cues. Each of these groups is assigned to a set of repair or re-engagement actions, respectively, executable by the robot, which range from simply stretching and breathing deliberately if the child seems to be tired, to standing up and doing squats or moving the arms if the child seems to have excess energy. Furthermore, most of these re-engagement actions invite the child to actively participate, except if the child appears only slightly

Dis-engagement Group	Behavioral Cues	Repair Actions
Tired	Rub eyes, Yawn, Support head with hands	Joint breathing, Stretching
Heightened activity	Move position, Undirected finger-tapping	Stand up and squat, Stand up and lift arms
Distraction low	Gaze away (from robot & tablet), Grimace	Wave, Whisper
Distraction high	Turn away, Move head from left to right	Chicken dance

Table 6.4: Overview of dis-engagement groups, the related cues and repair actions for training the WOz (reprinted with permission from Schodde et al. (2019)).

distracted, e.g., by grimacing or gazing away, and the robot just waves with its arm and/or whispers to them to regain their attention. However, if the child displays strong signs of distraction, e.g., by turning away completely or moving away from the interaction, and simpler re-engagement actions do not show any effect, the robot tries to re-engage the child by motivating them to join a dance as a last resort. This type of action is designed to channel the child's attention towards a different task while trying to keep them in a social interaction with the robot. Subsequently, the robot tries to guide their attention back to the vocabulary learning task with the assumption that their engagement is restored.

6.3.1.4 Explanation Verbalization

Equally to the previous studies, the system's adaptive tutoring behavior is based on the A-BKT model (see Chapter 5 for more details), which continuously tracks the child's proficiency level per target vocabulary and adapts the interaction accordingly. To allow for the SARTS to open up and verbally disclose this knowledge (open learner model), as well as the resulting consequences for the interaction flow, the underlying proficiency beliefs are divided into four discrete levels:

1. very low proficiency (skill belief: ∼0-25%), very easy task:
   e.g., *"For practice, I'm selecting an animal which I believe you don't know yet. To keep it simple, only three different animals stand for selection."*

2. low proficiency (skill belief: ∼25-50%), easy task:
   e.g., *"I'm selecting an animal that I believe you are not completely certain about yet. Try and find it among the three animals of the same color."*

3. medium proficiency (skill belief: ∼50-75%), medium hard task:
   e.g., *"Next, we are repeating an animal that I am quite sure you know. To make it a little more challenging, you have to choose from six animals."*

4. high proficiency (skill belief: ∼75-100%), hard task:
   e.g., *"I'm now selecting an animal that I am very certain you know. To strengthen your knowledge, it is hidden between eight other animals. I am curious to see whether you find it."*

These sentences are then followed by *"I spy with my little eye something that is a **[target word]**"*.

In addition, the SARTS also introduces the four dis-engagement groups that are tracked by the WOz and makes them explicit:

1. tired: *"I am under the impression that you are a little tired."*
2. activity: *"I think you cannot concentrate anymore."*
3. distraction low: *"I have the feeling that you are a bit distracted."*
4. distraction high: *"I think that you are too distracted to continue learning right now."*

While in both conditions the child is invited to join most of the following dedicated activities (except actions associated with "distraction low"), these introductory phrases that match the respective re-engagement behavior are only provided in the condition with explanations.

Overall, the explanations are used to reveal the system's perception of the children's performance, engagement and the resulting consequences for the task difficulty and course of interaction. This is intended to give the children a better understanding of their own evolving mental and knowledge state and, thus, to provide a scaffold for the perceived learning dimension to enable them to master the task they are faced with.

6.3.2 Measurements

In the following the new, adapted and more complex measurements applied in this study are discussed in detail. Since this study aims for addressing more than the children's cognitive learning, new tests for measuring their perceived learning and engagement are developed. Further, the previously used pre- and post-tests are adapted to incorporate the latest findings.

6.3.2.1 Engagement

To measure the children's engagement multiple factors are considered. First of all, the frequency of dis-engagements is counted whenever the WOz triggers a re-engagement behavior. Second, since the children are instructed that it is possible to terminate the interaction at any given time and are explicitly asked whether they want to continue playing after a re-engagement attempt, the number of played rounds can serve as a further indicator for interaction engagement. However, a shorter interaction time, because of fewer rounds played with the SARTS, does not necessarily account for a lack of engagement, since there are diverse reasons for an early termination. (1) The child decides to quit after being asked whether they want to continue after a re-engagement attempt. (2) The WOz observes dis-engagement cues for the fourth time and triggers a re-engagement behavior. This is set as a threshold to prevent the children to get stuck in the interaction although they are totally dis-engaged, do not learn a lot and are to shy to quit the interaction. (3) The system decides that the child performs very well, already learned enough ($avg(P(S_i^t)) \geq 75\%$) and, therefore, ends the interaction. However, this triggers a recap session consisting of three further tasks addressing maximal three of the weakest skills. (4) The maximum of 30 rounds is reached in the game and, finally, (5) the child decides to quit early due to annoyance or anxiety without being asked explicitly in connection with a re-engagement action.

6.3.2.2 Learning Gain

Similar to study 2, children's learning gain is assessed by conducting a pre-, post- and retention-test, which, however, are modified based on previously gained knowledge (cf. Section 5.4.6). First, to further lower children's shyness and anxiety prior to the interaction, the pre-test to measure children's prior knowledge of the words to be learned is designed as a conversation with one experimenter and is structured as follows. First, all 9 animals are presented to the child on a sheet of paper in randomized order. To ensure that all animals are known the experimenter asks the child to name the respective animals in German. Subsequently, the animal names are verbally presented by the experimenter in the foreign language (English), while the child is requested to tap on the corresponding picture. The given answers are noted by the experimenter without any feedback about their correctness.

Furthermore, the post- and retention-tests to measure the knowledge increase of the target words subsequent to the tutoring interaction are adapted, too. While the post-test is conducted right after the teaching interaction, the retention-test takes place at least 1 week later. Both incorporate the same animal pictures, layout and procedure, but are integrated into a more meaningful context as before. Now, the child is requested to feed the animals they were playing with for so long, since they became hungry (and feed them again after a while). To maintain the interaction's consistency the robot is also used for the post- and retention-test, in which it selects an animal in randomized order and says its English name to the child. Subsequently, the child has to gather a virtual grape from a basket on the bottom right of the screen and to move it to the intended animal to feed it (see Figure 6.5d). Afterwards the child receives again only neutral feedback, e.g., *"Thanks. The next animal we will feed is **[target word]**"*, so that they do not know whether they answered correctly or not.

6.3.2.3 Perceived Learning

In order to evaluate whether the children gained a better understanding of their own knowledge state regarding the target words, they are again presented with printed pictures of all nine animals subsequent to each post- and retention-test. They are asked to sort them into four knowledge categories, each representing one of the discretized and verbalized knowledge levels of the system ($0-25\%$, $25-50\%$, $50-75\%$, $75-100\%$). To simplify this self-assessment test for the children, colored squares representing the different categories are used that are printed on a separate sheet of paper and verbally explained by the experimenter.

- **red**: *"I don't have a clue what the English name of this animal is."*
- **orange:** *"I'm uncertain what the English name of this animal is."*
- **yellow:** *"I'm rather sure what the English name of this animal is."*
- **green:** *"I definitely know the English name of this animal."*

After the procedure is explained, the experimenter hands over one card at a time to the child. Each card displays one of the nine animals addressed during the tutoring interaction. The child's task is to sort the cards into those categories that best fits her own knowledge estimation while being informed that she can put as many cards as she wishes to each category.

In addition, the children's estimated knowledge is compared to their given answers in the respective post-test. Since the test results are just binary (right or wrong), the categories are collapsed into high (green and yellow) and low perceived knowledge (orange and red).

6.3.2.4 Re-Engagement Success

To measure the success of re-engagement actions the children are explicitly asked whether they want to continue to learn with the robot after each re-engagement attempt. The respective decisions are interpreted as an indicator of a successful or unsuccessful, respectively, re-engagement action. Additionally, if a child quits early after multiple re-engagements, all attempts that resulted in a continuation are still counted as successful and just the last attempt as unsuccessful.

6.3.3 Participants

A total of $n = 49$ children (22 female, 27 male) in the age of 4-7 years from three different German kindergartens participated in the study. They are randomly assigned to one of the two study conditions, while trying to balance age and gender within the conditions. However, nine children had to be excluded due to incomplete datasets that can be attributed to system crashes ($n = 2$), early termination of the interaction due to anxiety or annoyance shown by the children ($n = 6$), and impairments in language development, which might have affected the overall communication with the robot ($n = 1$). This results in a final group of $n = 40$ children (20 female, 20 male), in the age of 4-7 ($M = 5.43$; $SD = 0.54$), whereof 22 children participated in the experimental condition with explanations and 18 in the control group.

6.3.4 Procedure

Preparation

A few weeks before the study, parents of children in the respective age group are informed about the goal, setting and course of the study. Further, they are asked to fill out an informed consent form to allow for their children to participate in the study. Moreover, they are encouraged to contact the experimenters in case of any ambiguities or questions.

Icebreaker

At least one week before the actual study, the experimenters visited the kindergartens to introduce the robot and themselves to the children in a group session. As before, this is meant to provide a chance for all children to get to know the robot in a safe and known environment and to mitigate their initial anxiety (cf. Fridin, 2014a; Vogt et al., 2017). The introduction is designed as a conversation between the children, the robot and the experimenters. After the experimenters introduced themselves, they tell a short background story of the robot. In the meantime, the robot remains static. Afterwards, the children are encouraged to describe the robot's body parts and to find similarities (e.g., arms and legs), as well as dissimilarities (e.g., face and fingers), between the robot and themselves. Subsequently, the children are able to wake the robot up by calling its name collectively. Then, the robot stands up, greets

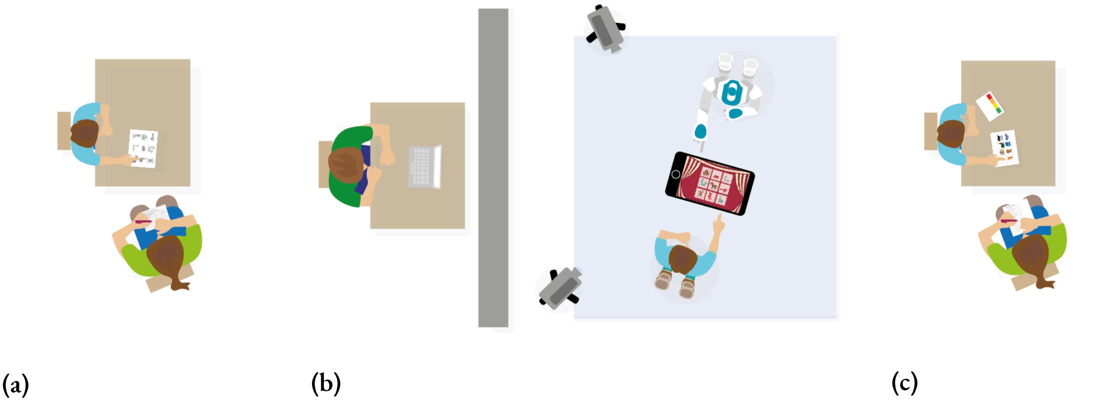


Figure 6.4: Experimental setup with (a) paper and pencil pre-test, (b) cHRI supervised by technician and (c) post-interaction interview (reprinted with permission from Schodde et al. (2019)).

the children, tells a few facts about itself and invites them to dance the “chicken dance” to loose its own tension. After dancing with each other, the robot goes back to a crouching position in order to rest, while the children are invited to touch or hug the robot carefully. Throughout the whole introduction session the children are free to ask as many questions as they like and when their first curiosity is satisfied, the experimenters and the robot say farewell.

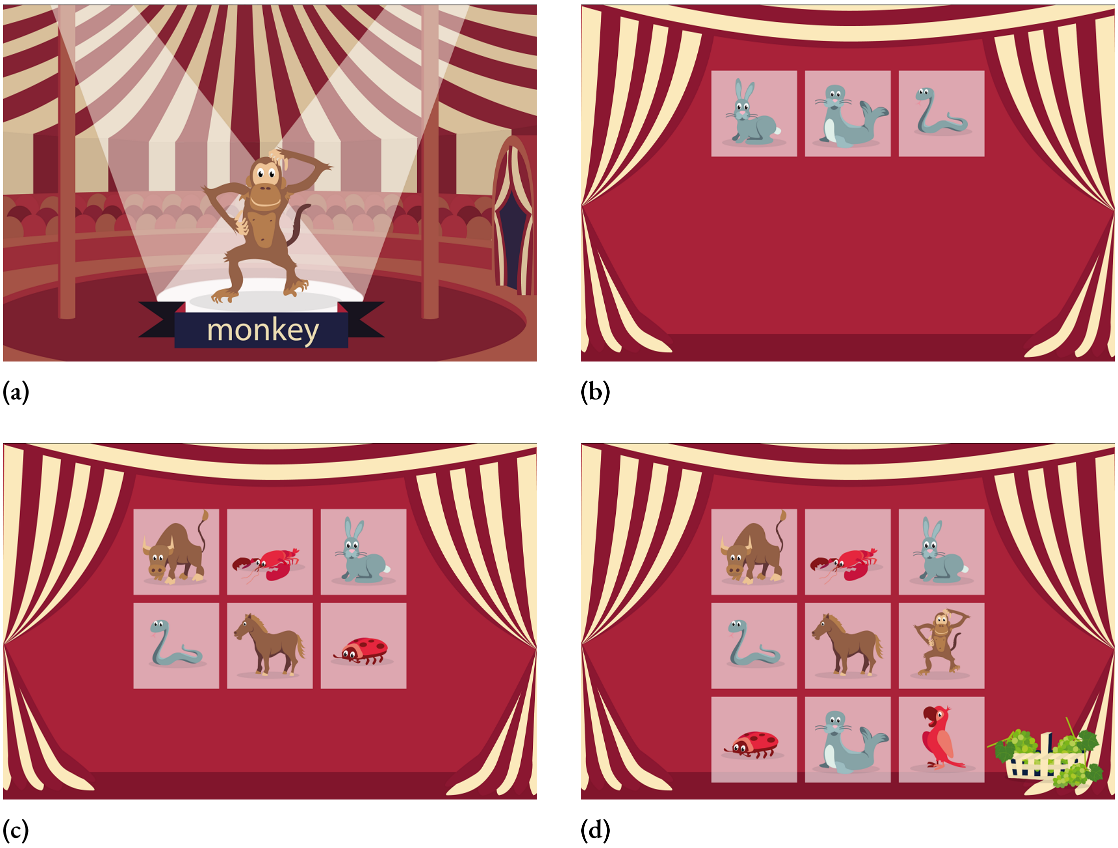
Pre-test

At the actual study day, the child enters the room either alone or with an educator and is asked to sit down at a small table together with one experimenter (see Figure 6.4a). Prior to the pre-test, the child is informed that she can stop at any time during the whole interaction. After the pre-test is completed, the experimenter guides each child to the main interaction spot and the child is asked to sit down on the ground in front of the tablet and the robot (see Figure 6.4b). Then, she gets a brief instruction on how to use the tablet and that she is about to learn English vocabulary while playing with the robot.

Tutoring interaction

If the child has no further questions, the interaction starts with the robot asking for the child’s name, age and previous knowledge of English vocabulary. Afterwards the circus is introduced, as well as all animals living in it. During this introduction the English name and a picture of the corresponding animal is displayed on the tablet screen (see Figure 6.5a). In the meantime, the robot pronounces the German and English animal name and executes an iconic gesture to support the *fast mapping* process. Subsequently, the child is asked to verbally repeat the English name to further strengthen the recall.

After all animals are introduced, the robot starts the real tutoring interaction and explains the “I spy with my little eye...” game. Then, two test rounds are played to verify that the child has understood all game principles before the actual language tutoring game starts. During the game each child plays 30 rounds, in which the A-BKT model chooses the skill to teach, here the English animal name, and a task difficulty expressed through a specific number of distractor animals displayed on the tablet screen (see Figures 6.5b & 6.5c). Then, the robot addresses the selected animal with its English name and the child has to tap on the corresponding animal on the screen.



(a) (b) (c) (d)

Figure 6.5: Screen captures from the learning interaction displayed on the tablet: (a) introduction of a new animal, (b) and (c) presentation of a set of animals for the task difficulty 1 and 3 during the training, (d) post-test including all target words (reprinted with permission from Schodde et al. (2019)).

Post-test and self-assessment of perceived knowledge

After finishing the game, either after 30 rounds or because of any other termination reason, except children quit early due to annoyance or anxiety (see Section 6.3.2 for the list of termination reasons), the post-test is conducted in which all 9 animals have to be fed (see Figure 6.5d). Then the robot explains that it is exhausted and needs to rest, thanks the child for playing with it and says goodbye. Afterwards the experimenter comes back and takes a picture of the child and the robot as a souvenir for the child, before the final interview takes place. For this, the child is asked to sit down at the small table again where she is interviewed about the interaction, as well as asked to estimate her perceived learning in a self-assessment test (see Figure 6.4c). Subsequently, the child is thanked and can go back into her class.

Retention-test

At least 1 week later, the post-test is repeated to test children's retention of all 9 animal words. On entering the room, the child is asked to sit down in front of the robot and tablet again and the robot starts the interaction. It explains that although it does not have enough time to play the game, the animals are hungry and have to be fed again. After the retention-test is completed, the robot thanks the child, says farewell and she goes back into her class.

Reasons for Termination	With Explanations (n = 22)	Without Explanations (n = 18)	Children in total
Played 30 rounds	11 (50%)	11 (61.11%)	22
Finished early due to high knowledge state (system's decision)	7 (31.82%)	5 (27.78%)	12
Finished early after re-engagement attempt (child's decision)	2 (9.1%)	2 (11.1%)	4
Finished early after 4 attempts to re-engage (system's decision)	2 (9.1%)	-	2

Table 6.5: Amount of children who terminated the interaction due to the listed reasons with respect to the experimental conditions (reprinted with permission from Schodde et al. (2019)).

6.3.5 Results

In the following the recorded data are analyzed with respect to the impact of explanations on understanding, motivation and learning gain. Further, the willingness to continue after states of dis-engagement is analyzed to investigate the effect of the robot's re-engagement attempts.

6.3.5.1 Children's Engagement

As described in Section 6.3.2, (1) the frequency of behavioral dis-engagement cues, such as yawning or heightened activity during the interaction (cf. Table 6.4), (2) the number of rounds played by the children, since they were free to quit at any given time, and (3) the reasons for termination are considered as indicators of children's engagement.

Based on findings of study 2 and 3 it was highly expected that children become easily dis-engaged after a couple of rounds due to the procedural repetition and their short attention spans (cf. David Cornish et al., 2009, p. 73). However, only twelve children displayed one or more behavioral cues of dis-engagement, whereof $n = 7$ played in the explanation condition and $n = 5$ in the control group without explanations. Furthermore, their shown cues were identical with those named in the empirical basis, e.g., they played with their clothes or stood up and sat down again in a different posture while not paying attention to the tutoring interaction. However, a comparison of the numbers of observed dis-engagement cues between the conditions did not reveal any significant difference.

This is unexpected since the tutoring interaction lasted more than 18 minutes in the control condition ($M = 18.16$ minutes; $SD = 2.18$ minutes) and 23 minutes in the explanation condition ($M = 23.76$ minutes, $SD = 3.97$ minutes) excluding the pre-, post- and self-assessment test for the perceived learning. This is even longer than in study 2 ($M = 18.63$ minutes; $SD = 3.05$ minutes) in which children's engagement significantly dropped towards the end (see Section 5.5). This means, although the interactions in the explanation condition are significantly longer ($t(38) = 5.36$; $p < 0.001$; $d = -1.75$), no significant higher rate of dis-engagement behaviors or early terminations can be observed. Furthermore, also the number of rounds played by the children revealed no significant difference (with explanations: $M = 26.55, SD = 5.37$; without explanations: $M = 27.67, SD = 3.58$). Although

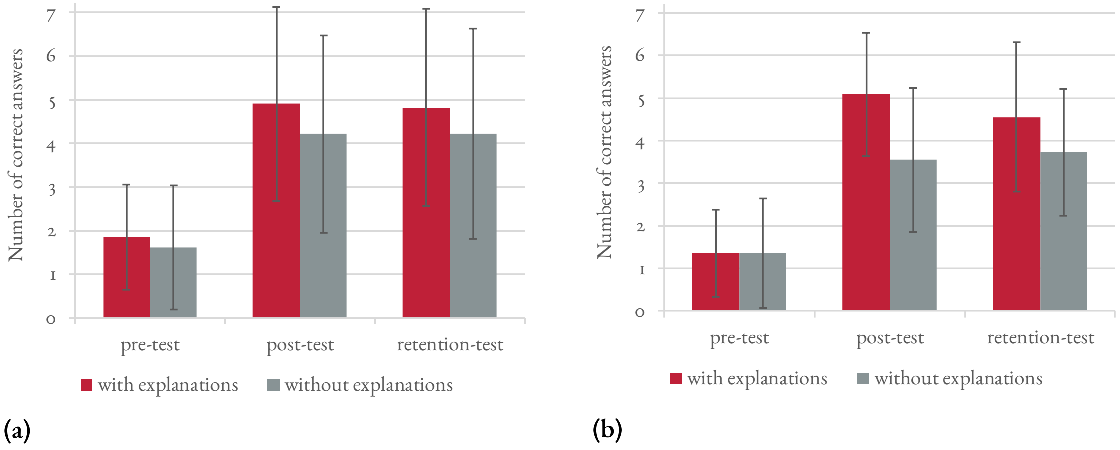


Figure 6.6: Average numbers of correct words in pre-, post-, and retention-test: (a) all children ($n = 40$) but with different interaction duration; (b) children who played all 30 rounds ($n = 22$; 11 in each condition) (reprinted with permission from Schodde et al. (2019))

a small tendency to play more rounds is observable in the control condition, which might hint to a higher engagement, the corresponding interactions were also much shorter on average. While this in turn could hint to a stronger engagement in the explanation condition, this is not supported when considering the numbers of early terminations.

As already mentioned, every interaction ended either because children aborted the interaction by themselves or after being asked subsequent to a re-engagement action, because the system decided to stop after four recognized states of dis-engagement or a high belief in the child's learned knowledge is attained, or because the maximum of 30 rounds is reached. As depicted in Table 6.5, only some children dismissed after re-engagement attempts, while the majority of children played the full 30 rounds or finished early due to they system's high belief of skill mastery. However, comparing the frequencies of terminations between the conditions does not reveal a significant difference, meaning, children continued or dismissed similarly often regardless of the robot's explanatory behavior.

In summary, these results do not support hypothesis H1, because children did not appear to be more engaged when interacting with the robot that explains its beliefs and decision-making.

6.3.5.2 Learning Gain

To investigate hypothesis H2, whether children learn more when the robot explains its beliefs and decision-making during the training stage of the interaction, a pre-, post- and retention-test was administered to measure the learned words before, immediately after and 1 to 4 weeks after the tutoring interaction. To ensure that the large time deviations for the retention-test, which occurred due to holidays in kindergartens, did not affect the retention-test results an independent samples t-test is calculated, which shows no significant difference of the average delays between conditions (with explanations: $M = 12.95$ days, $SD = 7.05$ days; without explanations: $M = 11.39$ days, $SD = 6.61$ days).

As depicted in Figure 6.6a, the average numbers of correct answers in all tests are higher in the condition with explanations as compared to the control condition.

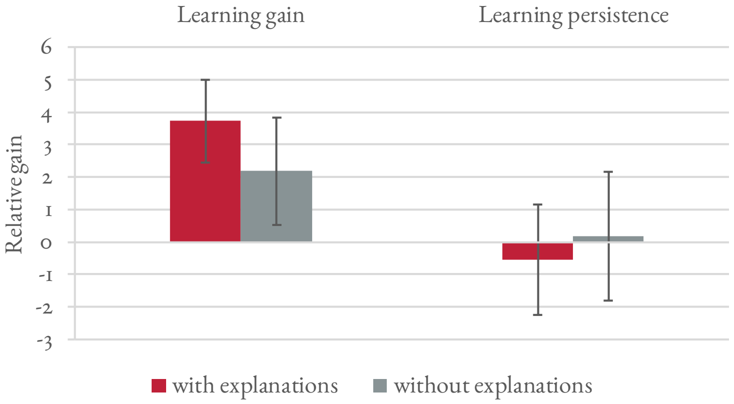


Figure 6.7: Average learning gain and persistence of children who played all 30 rounds ($n = 22$; 11 in each condition) (reprinted with permission from Schodde et al. (2019)).

Since large variances in the number of correct answer were already observable in the pre-test, the relative performance scores for the children's learning gain as the difference between the pre- and post-test, and learning persistence as the difference between the post- and retention-test are calculated. The comparison of these values between conditions revealed a higher learning gain in the with explanations condition ($M = 3.05, SD = 2.19$) as compared to the condition without explanations ($M = 2.61, SD = 2.00$). In contrast, the learning persistence is slightly higher in the control condition ($M = 0.00, SD = 1.82$) as compared to the condition with explanations ($M = -0.09, SD = 1.95$). However, comparing these results with an independent samples t-test revealed no significant difference for any of them.

Because of the high standard deviations observed during the analysis of correct answers, it is further investigated which possible causes might have affected children's learning. One reason for these high deviations might be based on the adaptive design due to which the termination points during the interaction varied from round 9 to round 30. Thus, in the following just children with an equal number of played rounds are compared in each analysis. In detail, just children with a number of 30 rounds ($n = 22$) are used and those are excluded who ended the interaction early due to dis-engagement ($n = 6$) or because of the system's decision that the child's skill mastery is high enough ($n = 12$). The resulting sub-sample of $n = 22$ children, with $n = 11$ per condition, is analyzed with an independent samples t-tests to compare the children's learning gains and persistence.

The corresponding results show that children who played 30 rounds within the explanation condition demonstrated a signification higher learning gain ($M = 3.73, SD = 1.27$) as compared to those who received no supportive explanations from the robot ($M = 2.18, SD = 1.66$; $t_{\text{one-tailed}}(20) = -2.45, p = .024, d_{\text{Cohen}} = -1.04$; see Figures 6.6b and 6.7). For the learning persistence, however, no significant difference between the conditions is observable (see Figure 6.7).

When reconsidering the small group of $n = 12$ children, who quit early due to the system's decision, a high learning gain from pre- to post-test in both conditions can be found (with explanations: $n = 7, M = 4.00, SD = 1.73$; without explanations: $n = 5, M = 3.80, SD = 2.78$). In the retention-test, however, the average diminished a bit in the explanation group ($M = -0.14, SD = 2.55$), while it

Categorization of target words	With Explanations (n = 16)	Without Explanations (n = 15)	Categorizations in total
(1) red ("no knowledge")	28	29	57
(2) orange ("uncertain")	15	26	41
(3) yellow ("rather sure")	30	23	53
(4) green ("good knowledge")	71	57	128

Table 6.6: Categorization of target words into the four given categories. The numbers represent the categorization frequencies combined over all children. Note that due to a break in the protocol the sorting task was only finished by 31 children (reprinted with permission from Schodde et al. (2019)).

remained the same on average in the control group ($M = 0.00, SD = 1.87$). Analyzing only those children who quit after a re-engagement attempt or because of four recognized dis-engagements revealed a negative learning gain between the pre- and post-test ($n = 4$, learning gain: $M = -0.50, SD = 1.29$), but a positive learning effect after a couple of weeks ($M = 1.25, SD = 0.96$) when playing with explanations. In contrast, children who played without explanations learned two new words from pre- to post-test ($n = 2.00, M = 2.00, SD = 0.00$), while they showed a negative learning persistence in the retention-test ($M = -1.00, SD = 0.00$). When just analyzing those children who got re-engaged by the robot and continued until the end of the interaction ($n = 6$) they showed a better overall learning gain (with explanations: $M = 4.00, SD = 1.00$; without explanations: $M = 2.33, SD = 2.52$) as compared to those who quit due to their dis-engagement. Further, with respect to their learning persistence, their knowledge slightly increased in the retention-test when playing without explanations ($M = 0.67, SD = 1.53$) and diminished a bit when explanations were provided ($M = -1.00, SD = 2.65$). However, because of the small sub-sample sizes and their uneven distribution between the conditions, no significance tests are calculated.

Summarized, hypothesis H2 that children will benefit from explanations about the robot's beliefs and decisions regarding their learning gains is not fully supported for all children. However, when considering only those who played 30 rounds and did not quit early due to high knowledge beliefs or dis-engagement, significantly higher learning gains are observable when explanations are provided as compared to the control group. This leads, at least, to a partial support of hypothesis H2.

6.3.5.3 Perceived Learning

To assess the children's perceived learning a self-assessment test was conducted in which they were asked to categorize the taught target words according to their low or high perceived knowledge (see Table 6.6). But due to a break in the protocol this test was only finished by $n = 31$ children.

Although most children seem to be fairly confident and categorized the target words into the green category ("good knowledge"), 57 out of 279 words are still categorized under the red category ("no knowledge"). Since the robot's explanations expressed the SARTS' belief about the child's skills, first, the impact of these explanations on the child's own estimated knowledge is tested. Therefore, the word-specific agreement between the child's self-assessment (high/low perceived knowledge) and the results in the post-test is calculated (right/wrong answer). An agreement is given, if children sorted the word into high perceived knowledge category and answered correctly or into the category of low perceived knowledge and answered wrongly (see Section 6.3.2 for more details).

However, the calculated overall agreement between children's self-assessment and the post-test results are nearly equal on average between conditions (with explanations: 64.58%, $SD = 14.28\%$; without explanations: 64.44%, $SD = 19.96\%$). Running the same analysis only for learners who played 30 rounds revealed that those in the explanation condition have a higher certainty on average in estimating their perceived learning (64.44% agreement; $SD = 15.56\%$; $n = 10$) than those in the control group (55.56% agreement; $SD = 19.25\%$; $n = 8$). In contrast, for children where the system decided to quite early due to a good performance, the difference between conditions diminishes again, although they show the best self-assessment accuracy on average (with explanations: 72.22%, $SD = 5.56\%$, $n = 4$; without explanations: 77.78%, $SD = 15.56\%$, $n = 5$). However, all found differences are not significant and, thus, hypothesis H3 is not confirmed.

6.3.5.4 Re-engagement after dis-engagement

To investigate whether a SAR can effectively re-engage children, who show cues of dis-engagement in a learning interaction (H4a), children's motivation to continue learning after a re-engagement attempt is analyzed. In total $n = 12$ children out of 40 showed 25 cues of dis-engagement. Half of these children ($n = 6$) were successfully re-engaged by the robot and continued the learning interaction until the very end (30 rounds). However, in one case the first attempt to re-engage was already unsuccessful. This child quit during the robot's re-engagement action and, thus, also skipped the post- and retention-test. In five other cases, dis-engagement re-occurred a second, a third or even a fourth time and resulted in termination of the interaction either by the child or by the study system. But still, half of the dis-engaged children could be re-engaged and motivated by the robot to continue the learning interaction and, thus, hypothesis H4a is, at least, partially supported.

To investigate whether the success of re-engagement attempts is affected by prior explanations why the system thinks a re-engagement action is necessary (H4b), further analyses are conducted. However, they revealed no significant differences between conditions. While 13 out of 17 (76.47%) re-engagement attempts were successful in the explanation condition, six out of eight (75%) yield a positive effect in the control group and, thus, hypothesis H4b is not supported.

In summary, although 30% of the children showed in sum 25 cues of dis-engagement and, thus, received re-engaging actions from the robot, 76% of these actions were successful. Consequently, most of the children who received a re-engagement action continued the learning interaction, whereas half of them even endured until the very end of 30 rounds. However, an effect for additional explanations by the robot was not observed.

6.3.6 Discussion

The presented study was aiming for evaluating the developed strategies to support young children's perceived and affective learning, as well as the engagement. While the scaffolding strategies to address the affective learning and engagement are used in both study conditions, the explanations to support learners' perceived learning are just used in one condition. As hypothesized, the language learning interaction provided by an adaptive SARTS resulted in higher vocabulary knowledge for all children independent of the study conditions. Furthermore, it can be observed that the knowledge learned from the SARTS persisted over several weeks.

Regarding the effects of explanations as a scaffold, the results revealed interesting facts. While the measured perceived learning and learning gain for the whole sample did not differ between the conditions, a more detailed follow up analysis revealed that the provided explanations did have a strong impact on a particular sub-group of participants, namely, children who completed all 30 rounds. These learners showed a tendency to a higher perceived learning in the self-assessment and achieved significantly better post-test results when the robot used explanations, compared to those in the control group. However, since these findings are a result of a follow up analysis with a reduced sample size of 11 children per group, it should be regarded with caution. But still, since the means for the whole sample of 40 children show a tendency into the hypothesized direction and the conducted tests on the sub-samples show a large effect size, this hints to the general existence of a difference caused by the robot's explanations and allows to make the assumption that the found effects will be even stronger with a larger sample size.

However, the results also revealed that the effect does not affect all children equally. For those children who quit the interaction earlier due to dis-engagement or a good performance no significant difference can be observed. The latter performed equally well with or without additional scaffolding by means of explanations provided by the robot, whereas the group of "low engagers" who quit early due to dis-engagement performed even worse if the interaction is lengthened by explanations. However, due to a small sub-sample size no reliable inferential statistics can be get and, thus, future work with larger sample sizes has to further explore the effects of explanations on children, who are rapidly dis-engaged or show a good performance during the interaction.

Since children, who ended earlier because of a good performance with a fast knowledge increase after a few rounds, showed a similar learning increase as compared to those who played 30 rounds with explanations they can be regarded as "fast learners". In contrast, those children, who take their time to get familiar with the learning content and, thus, played the full amount of 30 rounds can be regarded as "slow learners".

A possible explanation for the differing results between these groups and, in particular, for the positive impact on slow learners, is that the verbal explanations mitigated their uncertainty about their own performance, the robot's beliefs and the next tasks to be taken. In other words, in case of a non-optimal performance or increased uncertainty the SAR can provide explanations as a scaffold through which it reduces the felt uncertainty and the learner's cognitive load. However, although the used strategy affected the learning gain of slow learners in the expected directions, no significant impact on their perceived learning was found. This might be due to the observed problems of children to perform the self-assessment test to measure their perceived learning, which can be a result of their age group that just starts to develop the required self-monitoring abilities associated with SRL. Thus, they still might lack the ability to consciously access their own perceived learning, which might have negatively effected the self-assessment test results. However, the explanations still might have positively affected the perceived learning of all children, which is, at least, partially supported by the observed tendency towards better perceived learning assessment results for slow learners.

Fast learners, instead, may not have been able to benefit from the provided explanations at all, since they might have a higher task and learning proficiency in general, which prevented them from working in the ZPD so that no additional scaffolding from the robot was required. This is further supported by their higher prior knowledge already observable in the pre-test results, because of which they might have performed better right from the start and, thus, received more positive feedback throughout the interaction. This resulted in a higher cognitive and affective learning, which in turn may have resulted in an increased internal perceived learning. Because of this, the fast learners' uncertainty could have been further reduced, leading to an even smaller effect of the provided explanations and to comparable learning gains in both study conditions. Regarding the low engagers, however, it was observed that additional explanations seem to further distract them, which in turn resulted in a negative impact on their learning gains. The robot's longer speech acts might have bored them and, thus, they probably did not pay sufficient attention to the robot and the interaction to benefit from the provided explanations. This, in turn, could have resulted in the worse learning results as compared to those in the respective control group without explanations.

Regarding the provided scaffolding for the affective dimension and engagement a positive effect on young children's overall engagement and motivation was found. Although the interaction was longer as compared to study 2 (cf. Section 5.4.5), fewer cues of dis-engagement were observed. Consequently, the robot's default behavior including iconic gestures while introducing new words and an adaptive tutoring interaction combined with small socially supportive feedback behaviors, as well as re-engaging actions to repair dis-engagement, led to an overall more engaging interaction and, with that, to more concentrated and longer playing children. More specifically, the robot's re-engagement actions, such as waving or asking the child to stand up for stretching their arms, showed a success rate of 76%, so that 50% of the children who started to get dis-engaged continued the interaction up to the maximum number of 30 rounds after one or more re-engagement attempts by the robot. Those children further showed a good learning gain indicating that their dis-engagement was at least temporarily cured. Note that the robot directly asked the children after each re-engagement action whether they want to quit or continue the tutoring session. In general, this positive influence of re-engagement actions might be attributed to their alleviating effect on children's negative affective and cognitive states, such as inattentiveness or boredom. Especially physical actions, such as standing up or stretching the arms could have led to more concentration, as necessary for cognitive learning, and to higher engagement and affective learning. Furthermore, the robot's reactions to children's cues of dis-engagement unveiled its improved ability of actively keeping track of the learner. With this the robot appears not only to be able to keep track of children's knowledge state, but is also attentive to their affective states. Knowing that the robot can notice distraction or inattentiveness might have additionally fostered children's endeavor to concentrate on the task. However, making this ability even more salient by the use of verbal explanations did not increase this effect.

6.4 Summary

To allow for a SARTS to provide an optimal learning interaction for young children not only their cognitive learning has to be addressed but also their perceived and affective learning as well as their engagement. In particular, when trying to push the learner into the ZPD scaffolding is required to succeed within the tasks. However, the body of knowledge regarding which scaffolding strategies can be used by a SARTS to optimally support young children's learning in all its facets (**RQ2**) is still limited. Thus, this chapter aimed for extending this knowledge base by examining, inter alia, what are the relevant behavioral cues to track the engagement (**RQ2.1**), as well as which actions can be used by a SARTS to scaffold young children's engagement and affective learning during foreign language learning interactions (**RQ2.2**). To this end, an empirical basis was established based on expert's assessments and suggestions. It provides information about which engagement related affective and cognitive states are important for learning interactions and, thus, need to be tracked, how they can be tracked and how a SAR should react to regulate these particular states.

In addition, it was studied which strategies can be applied to scaffold young children's perceived learning during foreign language learning interactions (**RQ2.3**). With this goal in mind, a concept was developed to address the perceived learning dimension by allowing the robot to verbally express the SARTS' beliefs about the learner's knowledge and the resulting next steps in the learning interaction.

The identified strategies were evaluated in a study with $n = 40$ kindergarten children. As hypothesized, language learning with the developed SARTS led to higher vocabulary knowledge for all children independent of the study conditions. Furthermore, it was observed that the learned knowledge persists over several weeks. However, regarding the effect of explanations on young children's learning the results are inconclusive. While this scaffolding strategy showed a strong impact on the learning process of slow learners, fast learners' measured perceived learning and learning gains remained unaffected. This is probably due to fast learners' higher learning proficiency so that they did not need additional scaffolding from the SARTS. For slow learners, instead, this scaffolding seemed to be required. It might have mitigated their uncertainty about their own performance, the robot's beliefs and the next tasks to be taken, which then has positively influenced their learning performance.

With regard to the re-engagement strategies, the evaluation showed promising results. Most of the designed strategies were successful and half of the dis-engaged children had been re-engaged so that they played the game until the end and achieved similar learning gains as compared to those who were fully engaged throughout the whole interaction. However, for some children the strategies failed and might even have influenced their engagement negatively.

In conclusion, the different positive effects found in the presented study indicate that the presence of a SAR can be leveraged to maintain children's commitment to the learning task and also the interaction as a whole. These findings enrich the body of knowledge on how a SARTS can support the different dimensions of learning besides the cognitive domain and also point to new important questions to be answered in the future. In particular, the study results highlight again the differences between children and the importance of an adaptable SARTS to respond to their individual needs properly. However, this requires a more sophisticated *student model* containing a detailed student profile, which is not that easy to establish since just a small portion of the required information is directly accessible and the rest has to be inferred indirectly. Further, this information and the matching reactions (actions) have to be made available for the planning process, allowing for the SARTS to simulate their influences on the interaction and make an optimal decision.

In the following chapter a first step into the direction of a integrated model is taken that incorporates all findings of this thesis so far.

All students can learn and succeed, but not all on the same day in the same way.

— William G. Spady

7 An Integrated Model

At this point a suitable tutoring setting, structure and feedback behavior were identified and implemented in a basic SARTS. This system was extended by the A-BKT model to trace the learner's knowledge state (*student model*) and, based on this, to plan the next steps of the tutoring interaction with respect to teaching content and tasks difficulties (*pedagogical module*), which is able to increase children's learning performance and to maintain their engagement. In addition, the affective and cognitive states important for learning, as well as the related behavioral cues, have been identified, which can be used to recognize kindergarten children's engagement. Finally, different scaffolding strategies applicable by a SAR or SARTS, respectively, to address learning in all its dimensions have been identified and evaluated. However, these aspects are not fully integrated into the SARTS yet. Thus, this chapter focuses on how to combine them with the A-BKT to create an integrated model that is capable of considering the learning dimensions' interconnections and allows for an autonomous interaction with online adaptation (**RQ3**).

To investigate **RQ3**, this chapter describes how the previously developed A-BKT model can be extended to also incorporate the previously identified scaffolding strategies and their influences on the different learning dimensions. This includes influences of the learner's engagement, suitable scaffolding strategies, i.e., re-engagement actions and iconic gestures, as well as their costs and effects on the learning outcome. Furthermore, the robot's ability to provide explanations about its internal beliefs, e.g., the learner's skill mastery, and the consequences for the tutoring interaction is integrated (Section 7.1). Subsequently, the resulting integrated model is evaluated with simulations of different learner types derived from the previous studies, which allows to demonstrate the model's behavior during a tutoring interaction. Finally, the corresponding results are discussed and existing limitations are identified that have to be investigate in future studies (Section 7.2).

7.1 Decision-Theoretic Integration into the A-BKT Model

Usually the architectures of ITSs separate information about students' knowledge and engagement (*student model*) from the decision-making component (*pedagogical module*). However, combining both in a single model enables the system to simulate mutual influences of the strongly correlated learning dimensions, as well as the engagement, and to react accordingly. This in turn allows to avoid unnecessary behaviors, such as re-engagement attempts or iconic gestures, which would be triggered regularly as preventive actions if the decision-making module cannot assess the effect of the different learner states on the learning progress. Because this would lengthen the interaction unnecessarily, it might reverse the desired effects and cause, inter alia, boredom or frustration.

The previously developed A-BKT model already combines both modules and, thus, serves as a basis for the integrated model. It is based on the traditional BKT approach, which was extended and modified by adding an action decision node and splitting up the binary skill belief into six single states. These changes also necessitated the modification of the original update function of BKT. The new update function was carefully selected to maintain maximal flexibility with respect to further extensions of the model later on. Consequently, it is now easy to incorporate the learner's engagement, additional re-engagement actions, iconic gestures and tutoring explanations and, with that, to build the novel Probabilistic Tutoring Model for Autonomous Online Planning based on Predictive Decision-Making (ProTM). However, to be able to model all important influences of each scaffolding strategy and to simulate different learner types to demonstrate the behavior of the ProTM later on, information from further analyses of the previous studies is required.

7.1.1 Extended Analysis

To integrate the different scaffolding strategies two different pieces of information are needed, which can be derived from the previous studies. First of all, information about the influences of each strategy on the likelihood to observe a correct answer is required. This crucial information affects the action selection, in particular when supportive scaffolding actions are available for which the respective benefits need to be estimated and considered during the planning process. Second, the influences of the different strategies on the skill belief update need to be considered, meaning, whether a strategy supports the learner to achieve a higher performance in the dimension of cognitive learning or not. This is particularly important since such an influence would require the incorporation of the respective aspect into the model's belief update function. While the latter was already investigated in the respective studies, the influences of the different strategies on the likelihood of observing a correct answer were not analyzed yet. Thus, in the following this is done by examining children's frequency to answer correctly with respect to the effects of combining the A-BKT model with iconic gestures (study 2), re-engagement actions and explanations (study 4).

First, the effects of iconic gestures and the different tutoring strategies (adaptive versus random) on learners' frequency to answer correctly is analyzed by running multiple independent samples t-tests on data of study 2. Regarding the tutoring strategies, no significant effect is found. Children who played with the adaptive system achieved on average 70.00% ($SD = 13.51\%$; $n = 15$) correct answers during training compared to those in the control condition with an average of 74.17% ($SD = 13.31\%$; $n = 16$). For the use of iconic gestures, however, a significant effect can be found so that children who played with gestures answered significantly more tasks correctly on average than those playing without them (with gestures: $n = 16$; $M = 88.12\%, SD = 11.67\%$; without gestures: $n = 15$; $M = 70.00\%, SD = 13.51\%$; $t(29) = -4.005$; $p < .001$; $d = -1.439$).

Analyzing the effect of iconic gestures more deeply by examining the two hardest to remember words of each child, which were particularly addressed by the A-BKT model, revealed again a positive effect of iconic gestures. Children who had additional support through gestures from the robot tend to give less wrong answers to tasks addressing their hardest to remember words ($M = 2.27, SD = 2.26$) as compared to those playing only with the adaptive tutoring strategy ($M = 3.97, SD = 2.03$).

Second, analyzing the data of study 4 with respect to the effect of explanations on children's frequency to answer correctly revealed that those who played with explanations achieved a slightly lower frequency of correct answers on average ($n = 22; M = 72.09\%, SD = 14.95\%$) as compared to the control group without explanations ($n = 18; M = 73.82\%, SD = 10.31\%$). However, this difference is not significant. When comparing only children who played the full 30 rounds (slow learners; $n = 22; M = 69.96\%, SD = 8.34\%$) with those who quit early because of a high knowledge state (fast learners; $n = 12; M = 86.96\%, SD = 6.93\%$) this reveals a significant difference ($t(32) = -6.016, p < .001, d = -2.16$). This is not surprising since fast learners also showed a better learning performance because of which the learning interaction ended early. But when analyzing both groups with respect to the effect of explanations, no significant difference can be found anymore. While slow learners achieved an average of 69.03% ($SD = 10.59\%$; $n = 11$) of correct answers during the interaction with explanations and 70.85% ($SD = 5.69\%$; $n = 11$) without explanations, fast learners achieved an average of 87.45% ($SD = 7.59\%$; $n = 7$) correctly answered tasks with and 86.22% ($SD = 6.66\%$; $n = 5$) without explanations.

Finally, comparing the average frequencies of correct answers when children received re-engagement actions and quit early ($M = 54.86\%, SD = 6.86\%$; $n = 4$) with those who received re-engagement actions and played until the end ($M = 73.33\%, SD = 5.58\%$; $n = 6$) an obvious difference can be found. However, this difference diminishes when comparing the different conditions. Children who quit early showed an average correctness of 50.51% ($SD = 7.14\%$; $n = 2$) with explanations and 59.21% ($SD = 3.81\%$; $n = 2$) without them, while for those who played until the end the effect was even smaller (with explanations: $M = 72.22\%, SD = 7.70\%$; $n = 3$; without explanations: $M = 74.44\%, SD = 3.85\%$; $n = 3$). However, since this is just an analysis of observed subgroups in the whole sample and these groups are too small, no significance tests could be applied.

In summary, the average frequencies of correct answers provide indicators of how the different strategies influenced the likelihood to observe a correct answer during the interaction. Thus, the frequencies serve as basis for the following integration process of all strategies, as well as to inform simulations of different learner type, and will be regarded as the respective likelihoods of observing and providing a correct answer, respectively.

7.1.2 Model Extension

In the following, the results of study 2-4, as well as of the extended analysis above, are discussed with respect to the integration of the different scaffolding strategies (iconic gestures, explanations and re-engagement actions) into the A-BKT model. This includes results regarding the strategies' effects on the learner's learning progress, the likelihoods to observe a correct answer and the engagement.

7.1.2.1 Iconic Gestures

Some of the effects of iconic gestures were already examined in study 2, where they were continuously presented, since they were not yet integrated into the A-BKT model. Nonetheless, the results demonstrated that teaching actions supported by iconic gestures can boost the learning process when used together with an adaptive system (see Section 5.4). A possible explanation can be derived from the A-BKT model's skill selection behavior. It addresses those skills more often, which yield the highest answer error rates. Consequently, the children who learned with the A-BKT model got additional support or scaffolding, respectively, in form of iconic gestures for particularly those skills they were struggling most with. Moreover, this additional support further resulted in a significant increase of the likelihood to observe a correct answer from the learner (see Section 7.1.1) and, thus, in a better "flow" of the learning interaction, which in turn can increase students' perceived and cognitive learning.

However, because of these benefits on the students' learning process, the action selection of the adaptive model would prefer teaching actions supported by iconic gestures all the time, even if they are not required. As mentioned above, the positive effects shown by the A-BKT model combined with iconic gestures can be traced back to the individual support of the hardest to remember skills for each learner in the corresponding condition. Consequently, it can be assumed that iconic gestures are mainly beneficial for these particular skills and should be applied just for them. This prevents the interaction to be lengthened unnecessarily, which could cause negative effects on the learner's engagement. In fact, the results of study 2 demonstrated that although iconic gestures resulted in a higher overall engagement, a significant drop can be observed towards the end. For the adaptive system, however, this drop was barely existent (see Section 5.4). This allows the assumption that using the gestures only in situations where they are required and, thereby, not lengthening the interaction unnecessarily, might result in a smaller engagement drop and a significant interaction effect as it was observed for the students' learning progress. This is further supported by experts' suggestions in study 3 that providing a higher variability in the robot's behaviors can prevent children to dis-engage more frequently (see Section 6.2).

Consequently, to benefit from the positive effects of iconic gestures they are added to the A-BKT model as new teaching actions of the decision node TA^t, formerly known as A^t (see Figure 7.1 and Table 7.1). As before, each action represents one of four task difficulties, but is available in two versions now, either with or without iconic gesture support. Further, since TA^t still influences only the likelihood of observing a correct answer and the skill belief update, the positive influence of iconic gestures is modeled by modifying the respective values in the conditional probability tables. Finally, costs $C_{ta}(ta)$ for each teaching action $ta \in TA^t$ are introduced, in particular, for those with iconic gestures. These costs are offsetted against their usefulness in the respective teaching situation that is represented by the

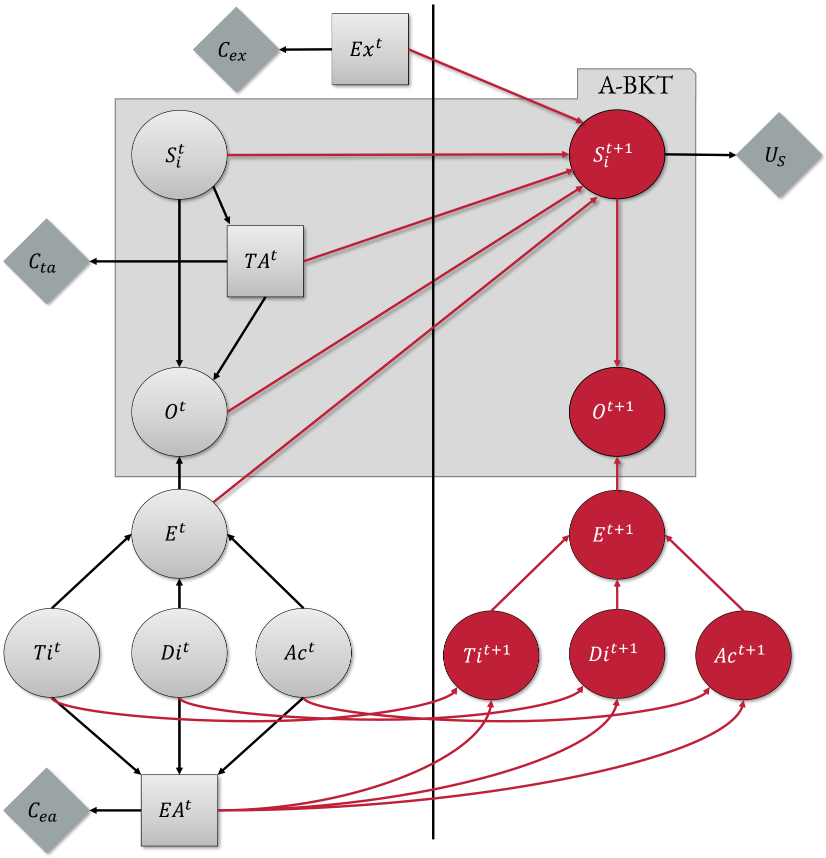


Figure 7.1: Visualization of the ProTM. The basic A-BKT (gray background) is extend with additional action nodes for explanations Ex^t and re-engagements EA^t, as well as the corresponding cost nodes C_{ex}, C_{ta} and C_{ea}. Furthermore, a cost for teaching actions TA^t is introduced and the learner's engagement E^t and the corresponding behavioral cues of being tired Ti^t, distracted Di^t or hyper-active Ac^t were added. Additionally, all new influences between time-slices are shown and their value is measured by a utility value U_L associated with the future belief about the children's skill mastery level.

individual error rate per target word (see Section 7.1.3). In sum, this results in applying iconic gestures only for hard to remember skills where the error rate is high enough, which is assumed to yield the best benefit for the learner.

7.1.2.2 Explanation Strategy

The explanation strategy is intended as a scaffold for learners' perceived learning and, with that, as indirect support for their cognitive learning. In fact, evidence for the latter was found in the results of study 4. Although children's frequency to answer correctly remained the same independent of the study condition (see Section 7.1.1), the results demonstrated that adding explanations can result in significantly higher learning gains (see Section 6.3). However, this effect was just observed for slower learning children. Already proficient learners, who went through the interaction quickly without many mistakes, did not benefit from them. Again, this highlights the possibility to adapt the interaction to each learner, which in turn can prevent possible negative effects on the learners' engagement, since these actions also lengthen the interaction.

Although, the results of study 4 did not show a significant higher rate of dis-engagement cues in the condition with explanations, this does not guarantee that no negative influences can occur. Since first indicators were already found that time consuming supportive strategies, such as iconic gestures, can have a negative influence on the learner's engagement, this might also apply to additional explanations. Perhaps, the interaction in study 4 was not long enough and applying them in a long-term study with multiple sessions might cause the negative effects to emerge measurably. However, this can be prevented by using also this scaffolding strategy adaptively, i.e., turning it on only if necessary, so that it results in a higher variability of behaviors shown by the robot, which, according to educators, can even support learners' engagement (see Section 6.2). Moreover, this maintains a productive interaction that is as short as possible and as long as necessary, which in turn might cause a higher long-term motivation of the learner so that she comes back and plays again later on. Finally, completing the learning content efficiently can again support the learner's feeling of flow and can lead to better learning outcomes (Craig et al., 2004; Hamari et al., 2016). Thus, the adaptive use of explanations can promote not only the learner's perceived learning but also her engagement, as well as affective and cognitive learning indirectly.

To benefit from these positive effects, a new action decision node Ex^t is added to the A-BKT model (see Figure 7.1 and Table 7.1). Furthermore, since study 4 demonstrated that adding explanations can result in significantly higher learning gains (see Section 6.3), while the likelihood to observe a correct answer remains the same (see Section 7.1.1), only the influence on the student's learning progress is modeled. Finally, costs $C_{ex}(ex)$ for applying the explanation strategy are added to the model, which are offsetted against the learner's global error rate. In contrast to the implementation of iconic gestures, this error rate is not skill-dependent and accumulates for all skills throughout the whole interaction. This enables the model to distinguish between the different learner types (slow and fast learners) and to apply explanations only if required. That is, if the error frequency is high enough (slow learner), the benefits of using explanations will be higher than their costs so that they will be applied by the robot. However, since no negative short-term effect was observed yet, this cost mechanic is only applied from the fifth round and the SARTS is forced to use explanations before. This is intended to provide optimal support for slow learners right from the beginning, while fast learners will stay unaffected.

7.1.2.3 Re-Engagement Actions

The learner's engagement is crucial for a learning interaction and can be influenced by the interaction itself, as well as by different scaffolding strategies. In general, a dis-engaged learner will probably not pay sufficient attention to the system or will not think deeply enough about the task, which can result in a higher likelihood of answering wrongly or even terminating the interaction. This was also observed in study 4, in which children who quit early due to four re-engagement attempts or decided to quit after such an attempt, which can both be interpreted as high dis-engagement, showed a higher rate of wrong answers as compared to the remaining children (see Section 7.1.1). Furthermore, they learned much less as compared to children who received re-engagement actions and continued until the interaction ended after 30 rounds (see Section 6.3).

Added Aspects	Model Modifications	Modeled Influences and Restrictions
Iconic gestures	new teaching actions for TA^t	Influence the belief update and the likelihood of observing a correct answer
	cost node $C_{ta}(ta)$	Restricts the use of gestures to hard to remember words
Explanations	actions decision node Ex^t	Influences the belief update
	cost node $C_{ex}(ex)$	Restricts the use of explanations to slow learners only
Engagement	latent variable E^t	Influences the belief update and the likelihood of observing a correct answer
	observable cues Ti^t, Ac^t, Di^t	Influence the Engagement E^t
Re-engagement actions	action decision node EA^t	Influences the single observable dis-engagement cues Ti^{t+1}, Ac^{t+1}, Di^{t+1}
	cost node $C_{ea}(ea)$	Restricts the use of re-engagement actions to a sufficient level of dis-engagement

Table 7.1: All modifications made to the A-BKT model to add the different scaffolding strategies.

But although the use of re-engagement actions can help learners to re-focus and re-engage into the interaction, they are also time consuming. Similar to the use of iconic gestures and explanations this lengthen the interaction, which might cause the contrasting effect of dis-engaging the children by boring them and, thus, they should also be used adaptively.

Consequently, to consider the learner's engagement during decision-making, E^t is added as a further observable variable, which influences the likelihood of observing a correct or wrong answer, as well as the skill belief update (see Table 7.1). As depicted in Figure 7.1, the learner's engagement E^t (discretized as low, mid and high) is calculated from three different cue groups for dis-engagement, i.e., tired (Ti), distraction (Di) and active (Ac). They can be regarded as abstract groups of behavioral cues identified in the expert interviews presented in Section 6.2. While the model can also receive the engagement from an external classifier, the direct inclusion of each single dis-engagement group yields further benefits. They can serve as a basis for the ProTM to select appropriate re-engagement actions to be executed by the SARTS. This is particularly important, since different dis-engagement reasons call for different re-engagement actions. For example, if a learner is very active and wants to move, an invitation to stand up and perform some easy physical exercises can reduce her urge to move and, subsequently, help to calmly sit down and concentrate on the learning interaction again. Therefore, another decision node EA^t is added that provides re-engagement actions for all possible combinations of dis-engagement groups, as well as for each group on its own. However, currently these actions are just placeholders that have to be mapped onto executable behaviors for the SARTS later on.

Finally, costs $C_{ea}(ea)$ for each re-engagement action $ea \in EA^t$ are introduced, which prevent the system to apply them instantly if the learner's engagement drops just a bit. Only if a moderate or high level of dis-engagement is observed by the tutoring system that drastically reduces the likelihood of observing a correct answer and, thus, hampers the learning progress, the cost of an additional re-engagement action is covered.

7.1.3 Formalization

After modifying the A-BKT model by adding all scaffolding strategies some of the equations need to be adapted, too. Since the different strategies showed influences on children's learning gain and/or frequency to answer correctly, the changes mainly apply to the skill belief update and the action selection, which will be described in the following.

7.1.3.1 Belief Update

The old A-BKT belief update is based on a simple Bayesian update function and, thus, can easily be extended to incorporate the novel extensions. Since all influences on the observable and latent variables are modeled as conditional probabilities, the learner's engagement E^t and explanations Ex^t have just to be added as further conditions. Adding teaching actions with iconic gesture support does not even involve an explicit change of the update equation, since they just have to be included into the action space TA^t and their influences on $p(O^t)$ and $p(S^{t+1})$ can be modeled via the respective conditional probability tables. In summary, the described adaptations for the belief update function result in the following equation applied to all $s_k \in S_i^{t+1}$:

$$\begin{aligned} p(s_k) &:= p(s_k|o^t, ta^t, E^t, ex^t) && (7.1)\\ &= \sum_{s_j \in S_i^t} \sum_{e_l \in E^t} \frac{p(s_k, o^t, ta^t, e_l, ex^t, s_j)}{p(o^t, ta^t, e_l, ex^t)} && (7.2)\\ &= \sum_{s_j \in S_i^t} \sum_{e_l \in E^t} \frac{p(o^t|s_j, ta^t, e_l) \cdot p(ta^t|s_j) \cdot p(s_j) \cdot p(e_l) \cdot p(ex^t) \cdot p(t)}{p(o^t, ta^t, e_l, ex^t)} && (7.3) \end{aligned}$$

where

$$\begin{aligned} p(t) &= p(s_k|s_j, o^t, ta^t, e_l, ex^t) && (7.4)\\ p(e_l) &= \sum_{ti \in Ti^t} \sum_{di \in Di^t} \sum_{ac \in Ac^t} p(e_l|ti, di, ac) \cdot p(ti) \cdot p(di) \cdot p(ac), \forall e_l \in E^t && (7.5) \end{aligned}$$

As previously mentioned, and also depicted in Equation 7.5, the ProTM calculates the engagement from three abstract cue groups, called tired (Ti^t), distracted (Di^t) and active (Ac^t). Each of these groups is associated with a set of behavioral cues (see Section 6.3.1.3 for more details) that have been identified in the conducted expert interviews (see study 3 in Section 6.2) and yield the potential either to be tracked automatically or to be wizarded by a human. While it is also possible to fed in the engagement level directly from an external classification framework, such as Affectiva Affdex (McDuff et al., 2016), this would restrict the selection of suitable re-engagement actions (see Section 7.1.2.3).

7.1.3.2 Predictive Decision Making

The process of predictive decision-making is divided into two major parts. First, the next skill to address is chosen and, second, the actions to be performed by the SARTS are selected. While the old A-BKT model had to choose just one action at a time, now, the action space is extended to also incorporate the new scaffolding strategies, which require to make multi-step decisions. Although a teaching action

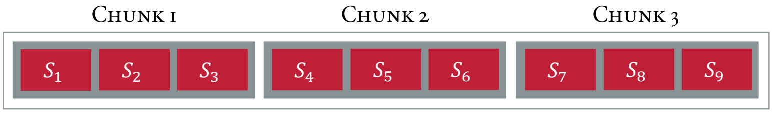


Figure 7.2: Visualization of skill space chunking with nine skills (red) distributed onto three chunks (gray) of length three.

is still chosen in each round, the learner might require further support through scaffolding, e.g., by executing a re-engagement action before a teaching action. This, in consequence, can result in a series of actions to be executed by the SARTS.

Choosing the next skill to address

Despite all the changes made to the A-BKT model to incorporate the novel strategies the basic skill selection remains the same. That is, the next skill to address is still identified by comparing the belief state $p(S_i^t)$ of each skill with the desired goal belief state $p(S_{opt})$.

$$next\_skill = \underset{S_i^t \in \mathbb{S}}{\operatorname{argmin}} \left[\alpha(S_i^t) \cdot KLD(p(S_i^t), p(S_{opt})) \right] \tag{7.6}$$

Since the application of the Kullback-Leibner divergence (KLD) requires at least some probability mass in each bin, $p(S_{opt})$ still represents the probability distribution in which $\approx 99.999\%$ of the probability mass is located in the last bin/state, meaning, being $\approx 99.999\%$ sure that the learner has mastered the current skill to 100%. The comparison is still done by incorporating the KLD, while the parameter $\alpha(S_i^t)$ is used to control the skill repetitions. It can range from 0.0 to 1.0, is decreased by 0.3 each time the same skill S_i is addressed and increased again by 0.15 if another skill is being practiced. This prevents the system to repeat a hard to remember skill over and over again and allows to switch to other skills in between before addressing the hard one again.

However, for large skill spaces this algorithm for skill selection becomes intractable. To resolve this problem, the skill space $\mathbb{S}$ can be divided into smaller parts (chunks) $\mathbb{S}_i$ of custom size n_{chunk} that can be taught one after another (see Figure 7.2). Thus, after all skills of a chunk are learned, the SARTS can switch to the next chunk and continue until all chunks or skills, respectively, are learned. This can be achieved by utilizing one of the termination reasons introduced in study 4. If the average skill mastery of all skills contained in the current chunk $\mathbb{S}_i$ is $\geq \varepsilon$, the system assumes that this chunk is mastered and switches to the next one.

Choosing the next tutoring and scaffolding actions

In contrast to the skill selection, the algorithm for choosing the next action to be performed in the learning interaction has to be modified. Up to now, the process of choosing the next action was limited to one action at a time, which was done for simplicity reasons. Further, since all actions had the same costs, the definition of suitable cost and utility values was postponed, because there was no need for a full fledged utility based selection function yet. However, the now added actions require to be able to make multi-step decisions and further require the consideration of different costs of actions dependent on the current situation. Therefore, the costs $C_{ta}(ta)$, $C_{ea}(ea)$ and $C_{ex}(ex)$ for each of the decision nodes *TA*, *EA* and *Ex* are introduced. All cost functions have their own dynamics and reduce the utility $U_s(s_j)$, which values the learning progress itself, so that the whole process is a classical cost-benefit analysis with the goal to find a series of actions to maximize the expected utility $EU(ta, ea, ex)$.

$$next\_actions = \underset{\substack{ea \in EA^t,\ ta \in TA^t, \\ ex \in Ex^t}}{\text{argmax}} \left[EU(ta, ea, ex) \right] \tag{7.7}$$

The expected utility *EU* is calculated based on classical decision theory and, thus, can be derived from the graphical representation of the ProTM:

$$EU(ta, ea, ex) = \sum_{s_k \in S_i^{t+1}} \sum_{o_j \in O^t} p(s_k | o_j, ta, E^{t+1}, ex) \cdot U_{all}(s_k, ta, ea, ex) \tag{7.8}$$

with

$$U_{all}(s_k, ta, ea, ex) = U_s(s_k) - C^*_{ta}(ta) - C_{ea}(ea) - C^*_{ex}(ex) \tag{7.9}$$

Furthermore, the engagement E^{t+1} is calculated based on the new dis-engagement cue values after an action $ea \in EA^t$ has been executed:

$$p(E^{t+1}) = \sum_{ti \in Ti^{t+1}} \sum_{di \in Di^{t+1}} \sum_{ac \in Ac^{t+1}} p(E^{t+1} | ti, di, ac) \cdot p(ti) \cdot p(di) \cdot p(ac) \tag{7.10}$$

while for each cue *CU*

$$p(CU^{t+1}) = p(CU^{t+1} | CU^t, ea) \cdot p(CU^t) \cdot p(ea) \tag{7.11}$$

$$p(ea) = \sum_{ti \in Ti^t} \sum_{di \in Di^t} \sum_{ac \in Ac^t} p(ea | ti, di, ac) \cdot p(ti) \cdot p(di) \cdot p(ac) \tag{7.12}$$

Thus, the utility U_{all} is calculated based on a set of actions $[ta, ea, ex]$ and the simulated effect of these actions on each state $s_k \in p(S_i^{t+1})$, as well as the engagement E^{t+1}. More precisely, U_{all} accumulates the costs for a specific set of actions with the utility U_s, which represents a value associated with the learning progress achieved by applying this set of actions. It is larger the closer these actions push the skill mastery belief $p(S_i^t)$ to the desired optimal belief $p(S_{opt})$.

Cost Functions

The dynamics of most cost functions depend not only on the corresponding action types but also on additional parameters associated with the respective field of application. For example, the costs $C^*_{ta}(ta)$ for teaching actions are calculated as follows:

$$C^*_{ta}(ta) = C_{ta}(ta) - \delta(S_i) \geq 0 \tag{7.13}$$

Although the former teaching actions still have no costs, they are required for the additional scaffolding through iconic gestures, since the system would prefer these actions over actions without gestures because of their benefits. To prevent this, the additional costs $C_{ta}(ta)$ are introduced, which are reduced by the skill dependent error rate $\delta(S_i)$, which ranges from 0.0 to 1.0 and represents the frequency of answering tasks wrongly that are associated with the skill S_i. Note that costs can never be negative so that a high $\delta(S_i)$ can reduce the costs of all novel actions maximally to zero, whereby costs for the former teaching actions remain constant. This results in converging costs of actions with and without iconic gestures until the benefits of using additional gestures predominate. Currently the value for $\delta(S_i)$ is calculated based on the last four given answers of tasks addressing S_i and the sweet spot at which the system will start to use iconic gestures is set to an error rate of 50%. This implementation is based on the extended analysis of the average error count for children's two hardest skills in study 2. While those who learned in the adaptive condition with iconic gesture support showed an average error count of $2.27(SD = 2.26)$, children who learned just with the adaptive system had an average error rate of $3.97(SD = 2.03)$. Consequently, iconic gestures are able to reduce the error rate per skill and it can be assumed that when two or more errors occurred, a hard to remember skill is taught and, thus, should be supported by iconic gestures to reduce the error rate of the learner for future tasks.

Similarly, also the costs of explanations $C_{ex}(ex)$ are reduced by the learner's error rate, but by her global error rate γ.

$$C^*_{ex}(ex) = C_{ex}(ex) - \gamma \geq 0 \tag{7.14}$$

This implementation is based on the observation of study 4 that children who learned slower and made more mistakes during the learning interaction benefited more from explanations provided by the SARTS. Consequently, the student's global error rate γ, which ranges from 0.0 to 1.0, is accumulated over the whole interaction and reduces the costs of providing additional scaffolding through explanations most when γ is maximal. Currently, the balance between costs and benefits is reached at an error rate of 23%. This is based on the extended analyses of study 4, which showed a global error rate of 30.04% on average ($SD = 8.34\%$) for slow learners compared to an average of 13.04% ($SD = 6.93\%$) for fast learners. Since in both groups no significant difference in the average error rates was found between conditions, it can be assumed that a global error rate above 19.97% (M + SD of fast learners) will identify slow learners. However, using an error rate right on the border increases the possibility of missclassifying a fast as a slow learner or the other way round and, thus, a threshold approximately in the middle of both averages is chosen.

In contrast, the costs $C_{ea}(ea)$ for re-engagement actions are independent of any external parameters. They are simply reduced by the positive effects of being highly engaged. For example, a dis-engaged child might fail at solving the presented tasks frequently, because of a lack of attention. In this case the likelihood $p(O = \text{correct}|E = \text{low}, ta, S_i)$ of observing a correct answer from the lowly engaged learner when using the teaching action ta for skill S_i is very low. Consequently, a correct answer is most likely the result of guessing instead of knowing the skill and the learning gain for the child will be small. Accepting higher costs, resulting from performing a re-engagement action beforehand and the teaching action afterwards, will improve the likelihood $p(O = \text{correct}|E, ta, S_i)$ and also the learning gain. The sweet spot, at which re-engagement actions become more beneficial than just trying to teach regardless of the learner's engagement, is currently reached at a belief of 70% for being tired or active, or 60%

for being distracted. Note that the threshold for distraction is slightly lower than the others, since it is the most crucial group currently implemented. Children who are distracted and, thus, pay no or less attention to the learning system are probably not able to hear the learning content or understand the tasks, respectively. However, combinations of different cues can cause the system to re-engage earlier. For example, if the ProTM recognizes a probability of 30% of being distracted and 40% of being tired, it will intervene and start a re-engagement attempt to address both dis-engagement groups at the same time. Note that the implemented cost functions and their respective parameters, in particular for the re-engagement interventions, are just based on educated guesses gained from study observations and, thus, need to be confirmed, refined or even replaced by empirically grounded values from future studies.

7.2 Simulation-based Evaluation

To demonstrate how a SARTS guided by the ProTM will behave in different situations with different learner types, a variety of simulations are carried out. As a basis, four different learner types are specified that learn with the SARTS and answer the provided tasks. They further simulate different dis-engagement behaviors, i.e., tired, active and distracted, that are used to calculate the respective engagement level, which then is infused into the ProTM.

7.2.1 General Mechanics of the Socially Assistive Robot Tutoring

To demonstrate the system's skill selection behavior, the learners have to learn six different skills during the simulated interaction, which is autonomously ended by the SARTS. This is done either after four re-engagement attempts or when the learner has reached an average skill mastery above $\varepsilon = 82\%$. Further, ε is used to control the system's chunking behavior, which, however, is just demonstrated in one of the simulations. In comparison to study 4, ε is raised to ensure that the learner mastered at least one task of the highest difficulty for each skill. This is not only intended to prove the learner's proficiency but also to quit the interaction before a real child might be bored by unnecessary repetitions of already learned content. Further, up to the threshold of 82%, the tasks are still challenging, since they just reached the highest difficulty so that a child can be assumed to stay engaged. Before ending the interaction a short recap is performed that contains three hard to remember skills. Although the interaction should be kept as short as possible, the recap is intended to check the learner's skill mastery, especially of those skills she struggled most with before ending the interaction. If this set contains less than three skills the recap lesson is filled up with skills randomly chosen from the remaining set.

7.2.2 Types of Learners

To demonstrate the behavior of the ProTM in different situations, simulations of multiple learner types are used. These types were observed in previous studies, which allows to base their answering behavior on the corresponding statistical values. That is, for each learner type, i.e., fast or slow learner, the likelihood of answering correctly is calculated with respect to the different scaffolding strategies. However, not all combinations of scaffolding strategies have been evaluated in studies yet. This particularly applies to combinations with iconic gestures so that the corresponding answer likelihoods are approximated. Since for children who learned with the support of iconic gestures the likelihood to observe a correct answer was 18.12% higher as compared to the controls, this value is treated as an offset for all groups of children learning with iconic gestures. For example, when a child usually has a likelihood of 50% to answer correctly, it will rise to 68.12% when iconic gestures are used. However, if the resulting likelihood reaches or exceeds 100% it is manually set to a slightly lower value, since a ceiling effect of iconic gestures can be assumed and a small likelihood of answering wrongly still remains.

Since also the A-BKT model's behavior to handle the influences of different engagement levels should be simulated, the learners' engagement behavior needs to be modeled, too. However, the previously conducted studies and interviews do not provide many indicators about how fast the level of engagement is falling on average or how fast dis-engagement cues are appearing, respectively. Consequently, the corresponding values describing the changes in learners' engagement are randomly picked from a predefined interval. This interval is intended to provide a meaningful basis to demonstrate the SARTS's behavior and is specified by a lower limit of 5% and an upper limit of 10%. This results in probability changes neither too small nor to big so that the system has to use re-engagement actions, but without the necessity to use them continuously. However, since the current model implementation calculates the learner's engagement level from observed behavioral cues, it is not modified directly, but rather through changes in the observed cue groups. Consequently, the randomly picked value describing the drop for the current round is randomly applied to one of these groups. This is done each round, except directly after a re-engagement action has been triggered, since it is assumed that the engagement level will stay for at least one round until the probabilities for the dis-engagement groups are rising and the learner's engagement is falling again.

Moreover, since the results of study 4 hints towards an impact of dis-engagement on children's likelihood of answering correctly, this is also modeled in the simulations. It can be assumed that children who quit early after four re-engagement attempts had a low overall engagement, while those who endured until the end can be assumed to have a higher engagement on average. The corresponding difference of 18.47% in the likelihoods of providing a correct answer is used to simulate the impact of dis-engagement on the different learner types. For example, for highly engaged slow learners the likelihood to answer correctly without explanations was 70.85%. If they would be lowly engaged the difference mentioned above is subtracted from the based likelihood associated with high engagement and, thus, the likelihood diminishes to 52.38%. Further, for learners with medium engagement a value right in between will be assumed, meaning, half of the difference will be subtracted again from the base likelihood, which results in a likelihood of 61.615% of providing a correct answer for slow learners.

In summary, based on the statistical results derived from the different studies and the definitions given above the following four learner types can be defined:

1. **Slow learner with constantly high engagement**: This learner is acting similar to a slow learner observed in study 4. The likelihoods of answering correctly are 69.03% with explanations, 87.15% with explanations and iconic gestures, 88.97% with iconic gestures only and 70.85% with neither of both.

2. **Fast learner with constantly high engagement**: This learner is acting similar to a fast learner observed in study 4. The likelihoods of answering correctly are 87.45% with explanations, 98.00% with explanations and iconic gestures, 98.00% with iconic gestures only and 86.22% with neither of both.

3. **Slow learner with drop in engagement**: For this learner the answering likelihoods of the first slow learner serve as baseline for a high engagement, which diminish with dropping engagement level. The latter is modeled by increasing the occurrence probability of a randomly chosen dis-engagement group in each round.

4. **Fast learner with drop in engagement**: For this learner the answering likelihoods of the first fast learner serve as baseline for a high engagement, which diminish with dropping engagement level. The latter is modeled by increasing the occurrence probability of a randomly chosen dis-engagement group in each round.

7.2.3 Results

In the following the simulation results of the four learner types learning with the ProTM are presented. To analyze the detailed behavior of the SARTS with respect to the chosen skills and actions based on the learner's knowledge, answering behavior and engagement level, all important interaction data are logged, analyzed and visualized in different graphs.

7.2.3.1 Slow and Fast Learner with Constantly High Engagement

Figure 7.3a exemplarily shows the system's behavior when confronted with a fast learning child with stable engagement. As depicted, the simulated learner makes just two mistakes and the tracked belief about her skill mastery is continuously rising. This good learning performance caused the SARTS to classify her as a fast learner and to turn off the explanations (gray background) after the fifth round, so that the interaction is not lengthened unnecessarily. Since the learner maintains a constantly high engagement, visualized as low probabilities for the behavioral dis-engagement groups, no re-engagement actions are triggered and the interaction ends after round 21.

Running the simulation multiple times ($n = 100$) reveals that fast learners need just 21.04 rounds on average ($SD = 2.74$) until the system ends the interaction because of a high knowledge state. Further, they need nearly no additional scaffolding so that the system applied 2.26 ($SD = 0.90$) iconic gestures for only 0.39 skills on average ($SD = 0.55$) and explanations only in 8.2 rounds on average ($SD = 5.50$). Note that in the first five rounds explanations are always used.

Analyzing an exemplary interaction with a slow learner, instead, reveals more interesting behaviors of the ProTM (see Figure 7.3b). Although this learner has also a constantly high engagement and, thus, the system shows again no re-engagement attempts, the decisions differ drastically with respect to teaching actions as compared to the fast learner. While the fast learner's interaction already ended after 21 rounds, the slow learner needs 37 rounds to learn the same amount of content, since she struggles more often and makes more mistakes. This, in consequence, results in the system recognizing her as a slow learner so that she is constantly supported by explanations. Moreover, this worse performance, especially for particular skills, further results in additional iconic gestures provided by the SARTS.

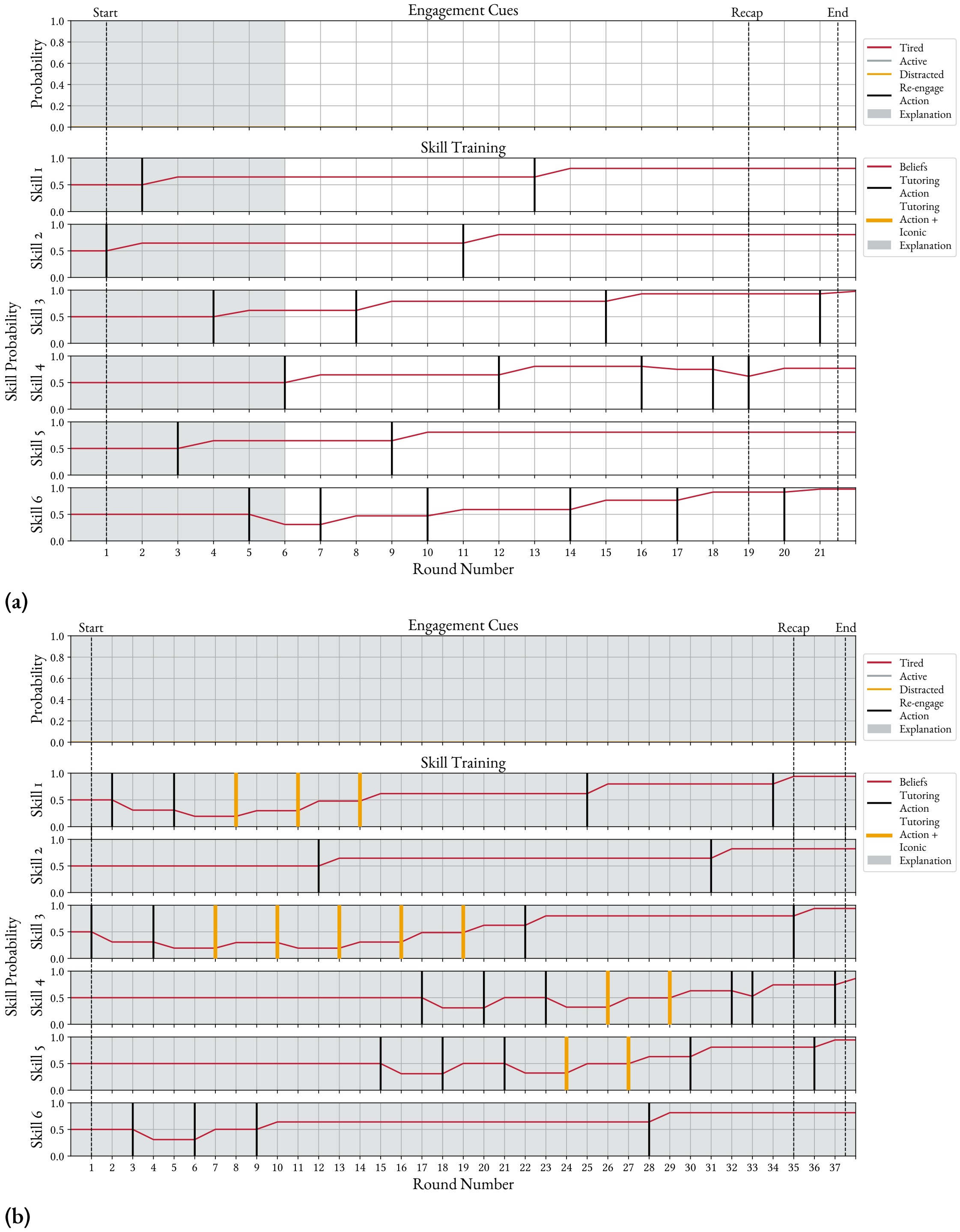


Figure 7.3: Interaction example for a fast (a) and a slow learner (b) with stable engagement on a high level. The topmost graph displays the probabilities of the different engagement cues per played round (here, all probabilities are zero). The six graphs beyond present the system's belief about the learner's mastery for the six skills to be learned. The vertical lines represent either normal teaching actions (black) or teaching actions including iconic gestures (orange). The gray background illustrates the use of additional explanations.

In detail, because the learner struggles in tasks addressing three of the six skills already in the beginning (Skill 1, 3 and 6), the system focuses on these skills first and tries to build a solid basis before introducing new content. For Skill 1 and 3 the ProTM even starts to provide iconic gestures (orange vertical lines) to support the learning process, since the learner makes several mistakes in a row. After tasks addressing these skills are answered correctly several times, the iconic gestures are withdrawn again, since a higher understanding of these skills is assumed.

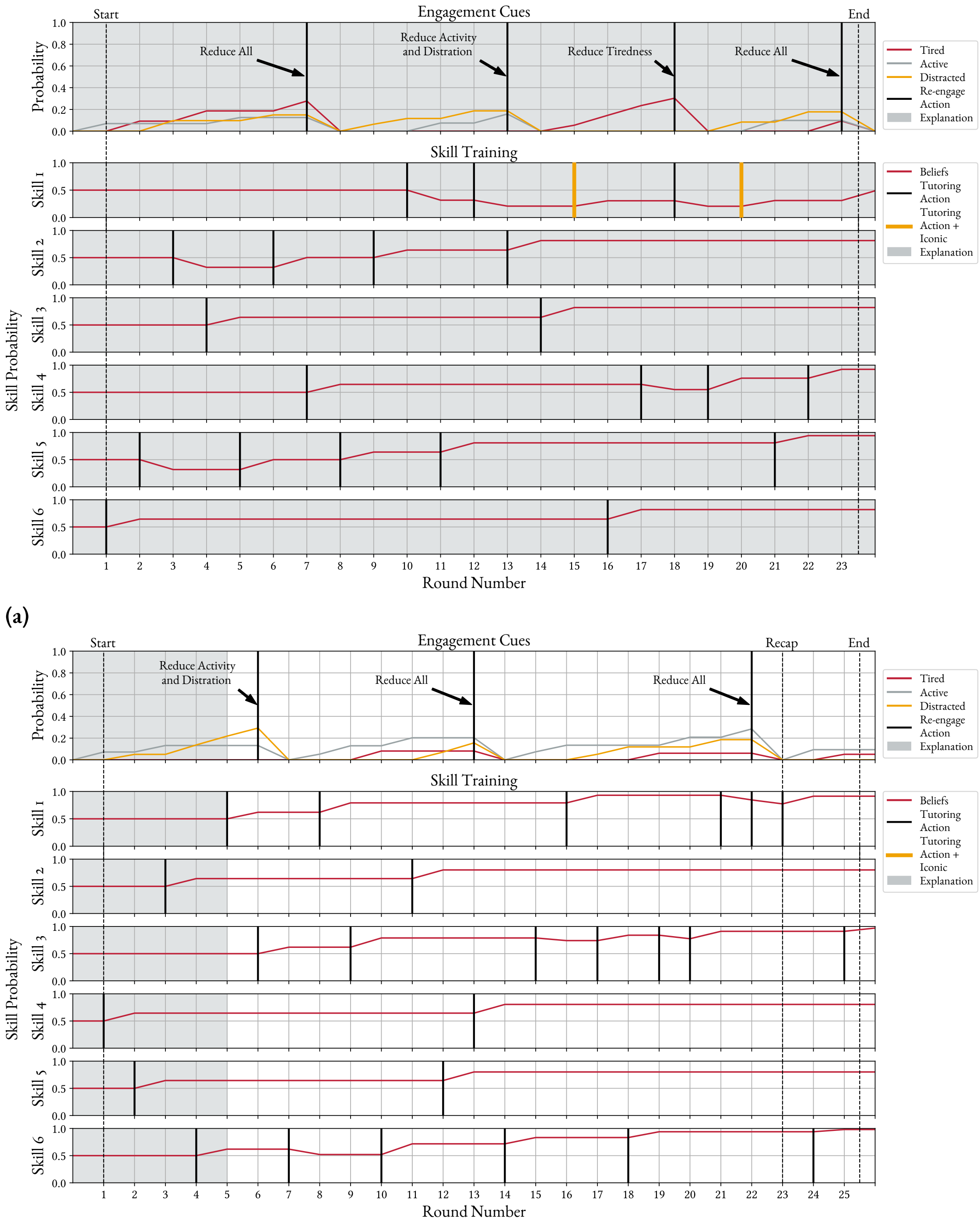
Regarding Skill 6, the learner seems to be fairly proficient after answering correctly two times in a row so that Skill 2 is introduced. However, this seems to be an easy one, because the learner answered correctly right from the beginning, and, thus, the remaining two skills (Skill 4 and 5) are introduced as well. Since a solid knowledge base for the initially introduced skills was build already, which is reflected by multiple correct answers of the learner in later stages, they are addressed less often and the system focuses on the novel ones. However, also the latest Skills 4 and 5 seem to be harder to remember for the learner and, thus, the system decides to support them with iconic gestures for two successive tasks until the learner improved her knowledge so that the gestures are withdrawn again. Towards the end all skills reached a similar belief level, so that the system repeats all of them until the termination condition is reached. Finally, the interaction is closed with a short recap including three of the learner's hardest to remember skills.

Running also this simulation multiple times ($n = 100$) reveals that slow learners need 32.95 rounds on average ($SD = 6.78$) until the system ends the interaction because of a high knowledge state. Furthermore, the system applies 4.51 ($SD = 3.11$) iconic gestures for 1.72 skills on average ($SD = 1.08$) and explanations for 27.87 rounds on average ($SD = 12.51$).

7.2.3.2 Slow and Fast Learner with Dropping Engagement

To demonstrate the ProTM's behavior with respect to handling the learner's dis-engagement, further simulations with slow and fast learners are carried out, while all of them show rising probabilities for the different dis-engagement cues.

As depicted in Figure 7.4a, the ProTM automatically chooses appropriate re-engagement actions for the slow learner to address the highest dis-engagement groups. This can either be an action to address all dis-engagement groups (re-engagement attempt 1 and 4), a combination of groups (re-engagement attempt 2) or just a single group (re-engagement attempt 3). As in study 4, the SARTS is configured to end after four re-engagement attempts, since this indicates that the learner struggles to concentrate in general and her engagement is dropping continuously. In this case, it is probably not reasonable to teach new content so that the interaction can be stopped for the moment to not harm the learner's general learning motivation. It can be continued at a later point at which the child might be less distracted, tired or active. However, the general teaching behavior of the system remains the same as before. It addresses the different skills, focuses on harder ones until they are learned and supports the learner with iconic gestures and explanations. Of course, the shown behavior of re-engaging the learner to maintain an appropriate learning atmosphere and to allow for the system to teach new content is independent of the learner type and, thus, is equally applied also for fast learners.



(a)

(b)

Figure 7.4: Interaction example for a slow (a) and a fast learner (b) with dropping engagement level. The topmost graph displays the probabilities of the different engagement cues per played round. The six graphs beyond present the system's belief about the learner's mastery for the six skills to be learned. The vertical lines represent either normal teaching actions (black) or teaching actions including iconic gestures (orange). The gray background illustrates the use of additional explanations.

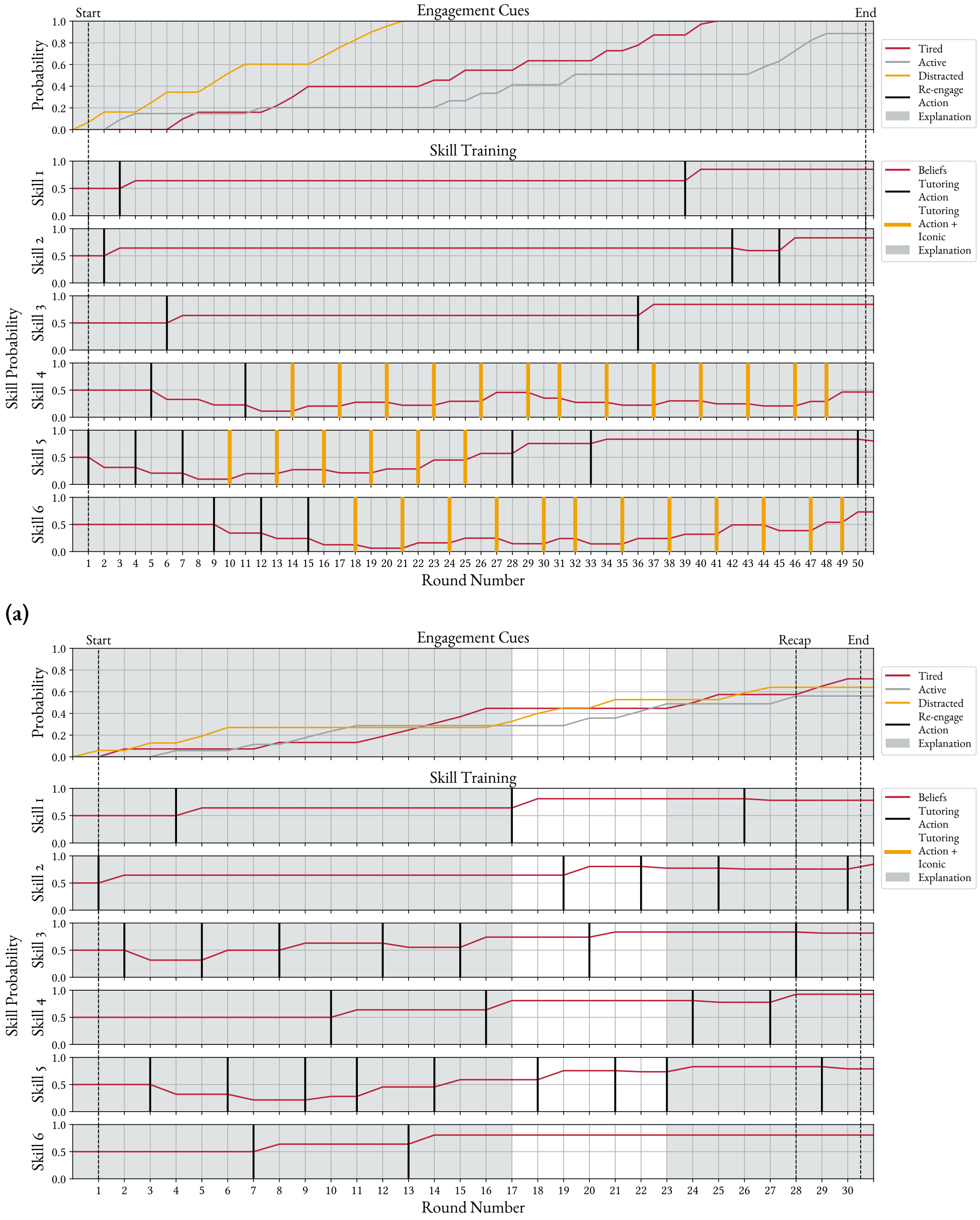
Although the fast learner's interaction is ended because she mastered all skills, she still received three re-engagement actions and needed four rounds more as compared to the fast learner with constantly high engagement (see Figure 7.4b). That is because the engagement influences the likelihood of answering correctly and the ProTM just intervenes when the influence becomes to strong.

Running both simulations multiple times ($n = 100$) reveals that for slow learners the interaction is ended after 24.63 rounds on average ($SD = 3.08$), mostly because of four re-engagement attempts ($M = 3.7$; $SD = 0.61$). Further, the system applies 2.49 ($SD = 1.48$) iconic gestures for 1.77 skills on average ($SD = 0.89$) and explanations in 20.94 rounds on average ($SD = 6.30$). For fast learners, instead, the interaction is ended after 23.17 rounds on average ($SD = 3.26$) because of a high knowledge state and they just need 2.45 re-engagements on average ($SD = 0.92$). In addition, they are supported with 1.95 ($SD = 0.98$) iconic gestures for 0.44 skills on average ($SD = 0.62$) and explanations for 10.94 rounds on average ($SD = 7.71$). Note that the small differences between the learners' average number of played rounds and iconic gesture counts is due to the different termination reasons. When running the slow learner simulations without the re-engagement termination condition, the results are more different. Now, the slow learners need 34.72 rounds on average ($SD = 8.94$) until the system ends the interaction because of a high knowledge state. Further, they received 5.2 re-engagement actions on average ($SD = 2.21$) and were supported with 4.63 ($SD = 3.79$) iconic gestures for 1.82 skills on average ($SD = 1.13$), as well as explanations for 30.75 rounds on average ($SD = 12.90$).

To demonstrate how the learning interaction may proceed if no re-engagement actions are triggered by the ProTM, two further simulations with both learner types are carried out that are limited to 50 rounds. In Figure 7.5a a worst case scenario for a slow learner is depicted. While the previously simulated slow learners performed fairly well throughout the interaction, this learner frequently struggles. Although the ProTM tries to support her with its remaining scaffolding possibilities by providing explanations continuously, adding iconic gestures for nearly all tasks addressing the skills 4, 5 and 6 and reducing the learning content temporarily to just these three skills, she does not manage to learn the whole content within the set rounds. This can be traced back to her lower likelihood of answering correctly, which is a result of her low engagement level.

Also for the fast learner a worse learning performance can be observed. Figure 7.5b shows that even a fast learner starts to struggle when she is highly dis-engagement so that the ProTM turns on additional explanations for the whole interaction time except for six rounds. Furthermore, although the fast learner managed to learn all skills within 30 rounds, comparing this with the performance of the fast learner supported by re-engagement actions reveals that the latter required five rounds less to finish the whole learning session including the recap.

Running also these simulations multiple times ($n = 100$) reveals that slow learners need 41.11 rounds on average ($SD = 8.11$) until the system terminates due to their high knowledge. Further, they are supported with 6.01 ($SD = 4.17$) iconic gestures for 2.15 skills on average ($SD = 1.03$) and explanations in 37.10 rounds on average ($SD = 11.97$). Fast learners, instead, reached this high knowledge state already after 27.00 rounds on average ($SD = 5.90$) and are supported with 3.33 (2.81) iconic gestures for 0.46 skills on average ($SD = 0.62$) and got explanations for 20.55 rounds on average ($SD = 9.72$).



(a)

(b)

Figure 7.5: Interaction example for a slow (a) and a fast learner (b) with dropping engagement level and no re-engagement actions. The interaction was limited to 50 rounds. The topmost graph displays the probabilities of the different engagement cues per played round. The six graphs beyond present the system's belief about learners' mastery for the six skills to be learned. The vertical lines represent either normal teaching actions (black) or teaching actions including iconic gestures (orange). The gray background illustrates the use of additional explanations.

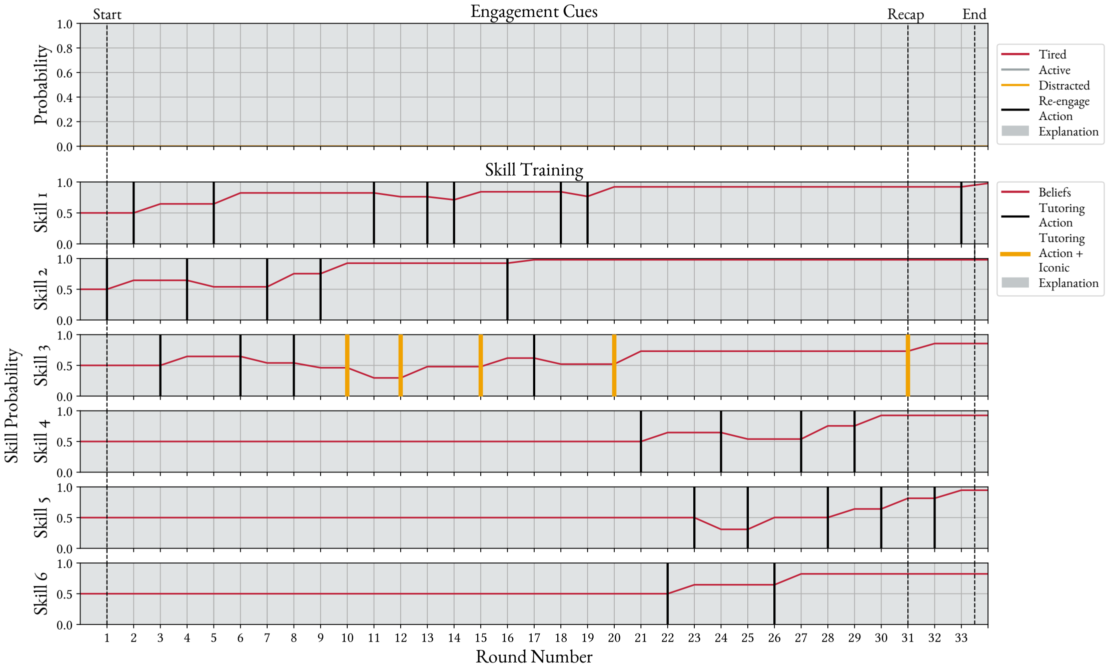


Figure 7.6: Interaction example for a slow learner with stable engagement level. The topmost graph displays the probabilities of the different engagement cues per played round. The six graphs beyond present the system's belief about the learner's mastery for the six skills to be learned. The vertical lines represent either normal teaching actions (black) or teaching actions including iconic gestures (orange). The gray background illustrates the use of additional explanations.

7.2.3.3 Skill Space Chunking

As mentioned before, the ProTM also supports skill space (learning content) chunking, which is depicted in Figure 7.6. To demonstrate this behavior a simulation with a constantly high engaged slow learner is carried out, while the model's internal chunk size n_{chunk} is set to three skills. As can be seen, this configuration lets the ProTM train just three skills at a time until a sufficient level of understanding is reached (belief $\geq$ 82%). Moreover, the system supports the learner with explanations, as well as iconic gestures, and automatically switches the chunk to train the remaining skills in round 21. After the second chunk also reached a sufficient level of understanding, the recap session is arranged with respect to the hard to remember skills over the whole learning content.

7.2.4 Discussion

The different simulations demonstrated how a tightly coupled, integrated system, which is enabled to reason about all possible actions and their effects to plan the next steps of a tutoring interaction, can behave. It can not only handle large skill spaces by autonomously splitting it into smaller chunks and switching between them, it can also adapt to different learner types easily and provide appropriate scaffolding to support each of them. For the fast learner with stable engagement, for instance, the learning interaction went smoothly, so that the ProTM just had to balance the attention between all skills, provided finally a short recap and then ended the learning interaction after the 21th round. In general, fast learners are very proficient learners and, thus, do not need much support during the learning process. This can also be observed in the behavior of the adaptive system, which turned off the explanations quickly after recognizing that they are not required.

Slow learners instead require more scaffolding throughout the interaction, also when they have a constantly high level of engagement. In general, a higher error rate can be observed for them during the learning interaction so that the ProTM provides supporting explanations more often or even throughout the whole interaction. Further, they can have "weak skills", meaning, skills that are harder to remember and require further support. Thus, the adaptive system is able to temporarily focus on those and to provide scaffolding through iconic gestures until a solid basis is established. Subsequently, the learning interaction can proceed as usual by maximizing the knowledge of all skills while balancing them under each other.

However, usually the ProTM also has to cope with a changing engagement level of the learner and needs to apply re-engagement strategies if necessary. This is intended to reduce the occurring dis-engagement indicators so that the learner stays concentrated and, with that, to increase her learning performance. In fact, the simulation results of a fast and slow learner with dropping engagement show that the adaptive system not only intervenes to provide re-engagement actions if the dis-engagement becomes too high but also chooses actions with respect to the shown dis-engagement groups, which allows to meaningfully support the learner within the respective situations. Still, while an overall highly engaged fast learner learned all six skills within 21 rounds, the drop in engagement results in a longer interaction with 25 rounds. Consequently, the learner's dis-engagement still influenced her learning progress even though the system provided re-engagement actions, which is explained by the way the costs for re-engagement actions are defined. Before a re-engagement attempt is started, a reasonable level of dis-engagement has to be observed to justify the resulting costs of lengthening the interaction. However, despite the four additional rounds, the ProTM still maintained a beneficial learning interaction for the fast learner, which is obvious when comparing her progress to a learning interaction in which no re-engagement strategies are used. Here, even the fast learner struggled more frequently. Although the adaptive system tried to support her by providing additional scaffolding by means of explanations, she still needed 30 rounds to finish the interaction, but, at least, she succeeded to learn the whole content.

For the slow learner, instead, this situation is even worse. While they performed reasonably well when the ProTM provided re-engagement actions, they seem to be totally overstrained with the interaction when their dis-engagement was not addressed. Although the adaptive system tried to support them as much as possible by continuously providing explanations and iconic gestures for a subset of skills that were identified as difficult, the learner would still need a large number of rounds to learn every skill, if even possible. Since the likelihood of answering correctly is dependent on the learner's engagement, the probability that she would manage to learn all six skills within a reasonable number of rounds is very low. Consequently, for the currently defined drop in engagement, it would be wise to provide re-engagement actions or even to stop the interaction and continue or restart it at another time, as done by the ProTM that was allowed to use re-engagement actions.

To summarize, the conducted simulations with different learner types provide a first impression of how the ProTM might work if applied in the wild. It successfully demonstrated its ability to autonomously react to the individual needs of different learner types. Although the simulations are only informed by incomplete information, they simulate a broad range of possible situations, which might occur during a tutoring interaction and need to be handled by a SARTS. However, young children usually provide an even higher variability in their abilities, especially regarding different skills, their engagement behavior, as well as reactions to different scaffolding actions, and the tutoring system still needs to react appropriately. Currently the simulations are just rough approximations of young children based on partial information about the learner types' answering behavior. Especially for the different strategies in combination with iconic gestures only few information are available so that the corresponding likelihoods to answer correctly are just approximated by educated guesses. Furthermore, literally no information about the average occurrence of the different dis-engagement groups can be derived from the previous studies. Thus, the simulation of learners' dis-engagement behavior is just modeled with regard to the creation of a meaningful use case to highlight the SARTS's general abilities to handle them. In addition, also the current model implementation is just based on parameters derived from previous studies that represent rough estimations regarding the time points when to apply iconic gestures, explanations and re-engagement actions to support the learner. Consequently, before a user study can be conducted that aims to validate the ProTM's behavior with real children, a variety of studies are required to confirm, refine or even replace the current model parameters. Further, more information about possible re-engagement actions is needed, in particular, about their effect on different dis-engagement cue groups so that they can be optimally reduced. Finally, the used thresholds for possible termination reasons are just based on educators comments or intuition. Although first validity indicators are already found for a small group of fast learners in study 4, who terminated early due to a very good performance, the thresholds need some further validation or refinement. This can be achieved through studies with bigger sample sizes to evaluate the already used parameters or empirical studies from which new or refined parameters can be derived.

In conclusion, the developed ProTM demonstrated promising results during the simulation-based evaluation, which, however, need to be verified in user studies with children in the future. Therefore, a more detailed knowledge base is needed that allows to refine or even replace the models internal parameters, as well as to derive further re-engagement actions appropriate for all different combinations of dis-engagement groups.

7.3 Summary

The goal of this chapter was to investigate how the identified scaffolding strategies, as well as the A-BKT model, can be combined into a single and novel approach that is capable of modeling the learning dimensions' interconnections and allows for an autonomous interaction with online adaptation (**RQ3**). In detail, this approach needs to allow for an interactive management of the tutoring interaction so that it is adapted to the learner's cognitive, affective and perceived learning, as well as her engagement.

To achieve this, the A-BKT model served as a basis, which was designed to be easily extendable. However, for integrating the different scaffolding strategies, further information was needed that, at least partially, could be derived from the studies 2 and 4. For this, further analyses were conducted, whose results were used to inform the integration process of the different strategies, as well as to derive simulations of different learner types to demonstrate the ProTM's behavior during a tutoring interaction. Overall four different learner types were defined to simulate a broad range of possible situations, which might occur and need to be handled by a SARTS during a tutoring interaction.

The simulation results provide a first impression of the models ability to actively manage the learner's engagement, learning speed and hard to remember skills by adapting the tutoring interaction autonomously, e.g., by applying appropriate scaffolding, if required. However, they still have to be validated in the wild with kindergarten children. Furthermore, the different parameters of the ProTM are currently defined through educated guesses derived from previous studies and might not be optimal for an interaction with kindergarten children, who will show a higher variability that might be more non-deterministic and, consequently, harder to handle by the adaptive system.

In summary, although further refinements of the ProTM's parameters are required, they provide a good starting point and the model already demonstrated its ability to actively manage the tutoring interaction and autonomously adapt it to the individual needs of the learner. It does not only manage the tutoring content and addresses the knowledge gaps to optimally support the learner's cognitive learning but does also provide appropriate scaffolding for the affective and perceived learning, as well as the engagement. Consequently, a SARTS based on the ProTM is able to provide a personalized tutoring interaction, which addresses learning in all its facets.

8
Conclusions

This chapter briefly summarizes the results, contributions and conclusions with respect to the research questions of this thesis (Section 8.1) and discusses its limitations, as well as the resulting future research directions (Section 8.2).

8.1 Summary and Contributions

This thesis contributes to the field of Intelligent Tutoring Systems (ITSs) or, more precisely to Socially Assistive Robot Tutoring Systems (SARTSs), with the major focus on extending the body of systematic knowledge on how a Socially Assistive Robot (SAR) can enrich a foreign language tutoring interaction for kindergarten children provided by a traditional ITS. Since this particular target group (children in the age of 4-6 years) has special needs and limitations, respective findings with older children probably do not generalize and, thus, the available information on how to extend an ITS with a SAR is still strongly limited. In addition, learning a language, as learning in general, is known to be a multidimensional problem so that this thesis focuses on the development of different approaches and strategies to address all of them, as well as learners' engagement.

Designing tutoring interactions for SARTSs

First of all, this thesis investigated the question which elements of language learning practices in German kindergartens can be implemented into a SARTS to provide a meaningful basis for foreign language learning interactions (**RQ0**). For this, observational recordings were carried out and analyzed with respect to transferable elements. The resulting empirical basis contains information about the general structure of language learning interactions, appropriate feedback behaviors and a tutoring game portable to a SARTS. Since it is derived from observations of experts' practice in kindergartens, it can be assumed that a SARTS, which follows the identified structure and rules, is designed appropriately for young kindergarten children.

Scaffolding learners' cognitive learning

Answering the second research question of how a SARTS can optimally address the cognitive learning of kindergarten children (**RQ1**) gets more complex and, thus, the problem is split into two sub-questions. In general, **RQ1** can addressed by adapting the tutoring content based on the learner's knowledge, e.g., by pushing her into the ZPD (cf. Vygotsky, 1978, p. 86). However, to achieve this the sub-question of how a SARTS can keep track of the individual knowledge and needs of kindergarten children to build up a *student model* (**RQ1.1**) has to be answered first. One widely used approach called Bayesian Knowledge Tracing (BKT) allows to tackle this question and, further, to investigate the second sub-question of how a SARTS can be enabled to select appropriate actions to adapt the interaction based on kindergarten children's individual knowledge state (**RQ1.2**). This is achieved by extending the traditional BKT with an action decision node so that the developed approach called Adaptive Bayesian Knowledge Tracing combines the *student model* and *pedagogical module*, which are both central parts of each ITS or SARTS, respectively (cf. Dede, 1986). This tightly coupled modeling further allows for impact simulations of all possible teaching actions onto the learner's knowledge and, thereby, enables a SARTS to choose the optimal teaching action in each state of the tutoring interaction so that it can be tailored to the individual needs of each student.

The evaluation with two user studies, one with adults and one with young kindergarten children, demonstrated that a SARTS can make use of the A-BKT model to autonomously support participants' progress in the dimension of cognitive learning. More precisely, learning with the adaptive model resulted in a better performance towards the end of the tutoring as compared to an interaction without personalization. However, this effect was not observable in learners' post-test results anymore. This is probably due to the limited number of interaction rounds during the study, because the adaptive system first focused on hard to remember words (see also Section 7.2), which resulted in less time for the learner to internalize the remaining target words. Furthermore, the tasks might have not been distinguishable enough in their difficulty so that "harder" tasks were still not challenging enough to allow for learners to internalize the words completely and to succeed in the post-tests. However, in general it can be concluded that the A-BKT model is able to keep track of learners' knowledge state and to identify weaker skills (**RQ1.1**), which are then addressed by the decision-making component with different task difficulties (**RQ1.2**). Although these results are not fully conclusive yet, they provide first steps towards answering the question of how a SARTS can optimally address the cognitive learning of kindergarten children during language learning interactions (**RQ1**).

Scaffolding learners' perceived and affect learning, as well as their engagement

In the second evaluation of the A-BKT model, with kindergarten children, another, unexpected and not intended but desirable effect was observed. The model was able to maintain children's engagement better as compared to a non-adaptive interaction. This can be regarded as a first step towards a SARTS that provides multidimensional learning support and leads to the next major research question about which scaffolding strategies can be used by a SARTS to optimally support perceived and affective learning, as well as the engagement of kindergarten children (**RQ2**). Again, this question can be split up into sub-questions, which can be tackled separately.

The first two sub-questions address the problems of which are the relevant affective and cognitive states, as well as, behavioral cues, to track the engagement of kindergarten children (**RQ2.1**) and which actions a SARTS can use to scaffold their engagement and affective learning (**RQ2.2**). Therefore, an empirical basis was developed together with educators from German kindergartens, which includes not only information about the important behavioral cues, as well as their interpretation with respect to affective and cognitive states, and engagement but also experts' suggestions about possible actions applicable by a SAR. This served as a basis for the implementation of different re-engagement actions, which were employed in the subsequent evaluation study. Additionally, a concept for training a human WOz was developed to achieve a high reliability for the engagement classification while avoiding the substantial costs of developing a robust classifier.

In contrast, scaffolding strategies to address the perceived learning of kindergarten children during foreign language learning interactions (**RQ2.3**), i.e., supporting their *self-evaluation* during learning, can be based on their skill mastery, as well as the interaction plan and history. This information is easily accessible, because the SARTS already trace this knowledge with the A-BKT model so that no further tracking approach is required. Based on this information an explanation strategy was developed in which the SAR reveals its beliefs about the learner's knowledge and the resulting consequences for the next steps in the tutoring interaction.

Both strategies, for addressing learners' affective and perceived learning, as well as their engagement, are evaluated with kindergarten children, which demonstrated their effectiveness in providing support during foreign language learning interactions. First, the previously identified dis-engagement cues were observed again, which can be regarded as a further indicator for their generalizability. Second, most of the re-engagement attempts allowed for the children to refocus and concentrate on the learning task again so that half of them were able to finish the maximum number of rounds. Consequently, these results further validate the identified behavior cues to track kindergarten children's engagement and the corresponding actions to scaffold their engagement and affective learning during foreign language learning and, with that, contribute to answer **RQ2.1** and **RQ2.2**.

The employed explanation strategy, however, only resulted in higher learning outcomes for slow learning children so that they were able to catch up with their more proficient and faster learning class-mates. Instead, for the latter who performed well and quit early due to high knowledge, the positive effect of explanations was not observed. But since fast learners already have a high proficiency level, the provided difficulty levels were probably not challenging enough, so that no additional scaffolding for their perceived learning was required. Furthermore, although these results hint towards the existence of a strong connection between perceived and cognitive learning, the direct influence on the former was not confirmed. As already mentioned, conducting tests with young kindergarten children is complicated (cf. Belpaeme et al., 2013a), which was also approved within this study. It was observed that many children struggled with the self-assessment test specifically designed to access their perceived learning, which in turn resulted in non conclusive results. In summary, these results provide a partial but still valuable contribution with respect to answering **RQ2.3** about which strategies can be applied to scaffold the perceived learning of kindergarten children during foreign language learning interactions.

Integrated model

Finally, this thesis investigated the question how to combine multiple strategies addressing the different learning dimensions and engagement into a single model that is capable of describing different action influences, as well as the learning dimensions' interconnections and allows for an autonomous interaction with online adaptation (**RQ3**). Since the A-BKT model was designed to be easily extendable and already proven to be able to support children's learning progress, as well as engagement, it served as a basis for answering this questions and, therefore, all identified scaffolding strategies were integrated. This enables the resulting Probabilistic Tutoring Model for Autonomous Online Planning based on Predictive Decision-Making (ProTM) to reason about the different information about the learner, i.e., knowledge state and level of engagement, as well as their influences to choose the next actions correspondingly. Since the integration process was guided by educated guesses derived from all previous studies of this thesis, which still need to be refined and confirmed with empirical studies, it was just evaluated by simulating different learner types to demonstrate the general behavior of the integrated system. The selected learner types were also observed in previous studies so that their answer behavior could be based on the respective statistical values. The results of these simulations nicely illustrate the ProTM's ability to identify the individual needs of each learner type and to tailor the interaction accordingly. However, this evaluation can not replace a user study with real children, who will probably provide a broader variety of individual differences and nondeterministic behaviors, which the SARTS has to handle. But still, the ProTM showed promising results with respect to the integration of different strategies to address all important aspects of learning and modeling their interconnections, which allows to find the best action sequence in each situation. Consequently, this can be regarded as a first step to answer **RQ3**.

The Overall Goal

Coming back to the overall question of whether and how a traditional ITS can be enriched by a SAR, the novel approaches, strategies and evaluation results presented in this thesis support the assumption that it is generally possible. But still, the way how the SAR is used by an ITS is particularly important, which is also highlighted by the latest study within the L2TOR project. While the robot guided the children through multiple sessions of a carefully designed tutoring interaction spread over several weeks and tried to support them with iconic gestures, no significant improvement in their learning gain with respect to the use of a robot compared to a tablet was found (Vogt et al., 2019). However, this study lacks one important aspect, namely, adaptivity. All sessions were scripted and each single target word was continuously supported by iconic gestures, which in turn lengthened the interaction significantly. As discussed before, this can harm learners' engagement and, with that, reduced their learning gains. Further, the missing individual support, in particular for children's knowledge gaps, lead to a default session comparable to traditional classroom instructions in schools, which was argued to be outperformed through individualized one-on-one tutoring interactions by at least 0.8 standard deviations on average with respect to the learning gain (VanLehn, 2011). Designing the interaction to be more flexible, instead, so that a SARTS can make use of the robot's benefits adaptively with respect to the individual needs of each learner, can bring up more positive effects of a SAR during such a language learning interaction. In fact, previous research already highlighted the usefulness of a robot during learning interaction (Han et al., 2005; Hyun et al., 2008; Kose-Bagci et al., 2009; Leyzberg et al., 2012) and demonstrated that it can increase students' learning gains up to 50% (Kennedy et al., 2015). However, caution is still advised when integrating a SAR into an ITS without sufficient knowledge about the effects of its actions. Kennedy et al. (2015), for instance, demonstrated that overdoing it by designing the robot to be too social and integrating too many behaviors distracts the learner and harms the whole learning process (Kennedy et al., 2015). To prevent such negative effects, this thesis enriched the body of knowledge on different strategies applicable by a SAR during language learning interactions and their respective effects on the different learning dimensions. These strategies can be combined with an adaptive teaching model to achieve a personalized one-on-one tutoring interaction that supports each learner with respect to her individual needs.

In conclusion, this thesis provides further support for the general feasibility of enriching a traditional ITS with a SAR. Moreover, the developed A-BKT model demonstrated its ability to support young kindergarten children in their cognitive learning and engagement, whereas the different scaffolding strategies showed their positive influence on the remaining dimensions, as well as the engagement. Finally, it was demonstrated how these strategies can be integrated into the A-BKT and its online decision-making process. This allows for a SARTS to apply these actions, adaptively, based on the learner's performance, knowledge and engagement state, and, with that, to provide an individualized interaction for each child. Consequently, this thesis provides new valuable insights for the respective field of research, but also points towards new important questions to be addressed in the future.

8.2 Limitations and Future Research Directions

Despite the mentioned contributions, this thesis still yields some limitations. Currently, the parameters of the A-BKT model, which, inter alia, specify the learning gain when a correct answer is provided, are manually set. Although they are informed by learning techniques, such as the *spaced repetition* system and the conducted evaluation studies validated that these initial parameters work in general, they are probably not best fitting for each individual learner. Thus, these parameters need to be refined and adapted during the interaction, e.g., by applying the Baum-Welch algorithm (Welch, 2003), for which, however, more data are required. To achieve this, a long-term interaction study can be conducted with multiple sessions spread over several weeks (cf. Leyzberg et al., 2018). Further, repeated sessions combined with new and harder task difficulties probably result in more meaningful post-test results, since the A-BKT model has more time to address the individual knowledge gaps of each learner, whereas harder tasks can be presented for the well known skills to internalize the attained knowledge.

In addition, although the presented studies implied or confirmed the benefits of the identified and implemented scaffolding strategies, their integration into a SARTS points towards new important questions. Study 2, for instance, demonstrated that iconic gestures in general are highly supportive when combined with the A-BKT model, which probably can be traced back to the individual support of weakest words particularly addressed by the adaptive interaction course. However, this assumption could not be validated yet and it is further mostly unknown when a word has to be classified as "hard to remember" so that iconic gestures should be applied. To resolve this, further studies have to be conducted that try to validate the benefits of an adaptive usage of iconic gestures, as well as to identify the sweet spot when the gestures should be provided as an additional scaffold for a particular word.

Regarding learners' engagement two different aspects require further attention. First, to allow for a SARTS to run fully autonomous, the identified behavioral cues need to be combined into a sophisticated engagement classifier. Therefore, further data recordings are required, which can be informed by the identified cues to narrow down the possible feature space. This will reduce the costs to establish a sufficient dataset by reducing the required data for training. Moreover, this classifier has to be evaluated in a user study with young kindergarten children to ensure its validity and applicability for this age group. After the engagement level can be classified reliably, further re-engagement actions for the SARTS have to be specified and validated. During the planning process they should be selected with respect to the behavioral cues or groups of cues used for the classification, since different dis-engagement cues/groups, such as tired, distracted or heightened activity, or combinations of those can account for different re-engagement actions. For example, a very active child might be re-engaged by providing some small physical exercises to reduce the excess energy so that she can concentrate on the learning interaction again. For a distracted child, instead, small verbal behaviors of the SAR can be sufficient, e.g., asking the child to direct her attention back to the interaction. This further demonstrates the robot's sensing capabilities, which can cause cautiousness in children so that they try to be more attentive and concentrated on the learning interaction (see Section 6.2.4).

Furthermore, the results for explanations as a scaffolding strategy need to be confirmed since the effects were found in a subsequent analysis of a sub-group of participants. This group was quite small and, thus, a replication of this study focusing on the identified sub-groups of fast and slow learners will be required to validate the general applicability of the findings. Additionally, the differences between both groups need more investigation regarding their integration into the ProTM. That is, it has to be clarified, at which point a fast learner turns into slow learner and the other way round. This will allow for an adaptive system to increase the precision in identifying students' learner type autonomously and based on this to turn on explanations as an additional scaffold if necessary.

After all this information is collected and infused into the ProTM, it can be used to investigate possible synergy effects of different strategies in different situations. For example, can explanations further increase the benefits of iconic gestures or does the accumulated lengthening effect of both significantly harm learners' engagement and concentration? These are also important aspects, which can be traced back to the tightly coupled learning dimensions and their influences on each other.

Finally, since the ProTM was only evaluated with simulations, the collectible information above will allow to validate its benefits in the wild with kindergarten children. This is preferably done within a long-term interaction over several sessions and weeks, to provide sufficient data for the underlying model to refine its parameters so that it can optimally address each individual learner.

References

Alavi, M., Marakas, G. M., and Yoo, Y. (2002). A comparative study of distributed learning environments on learning outcomes. *Information Systems Research*, 13(4), pp. 404–415. DOI:10.1287/isre.13.4.404.72.

Alemi, M., Meghdari, A., and Ghazisaedy, M. (2014). Employing humanoid robots for teaching english language in iranian junior high-schools. *International Journal of Humanoid Robotics*, 11(03). DOI:10.1142/S0219843614500224.

Alemi, M., Meghdari, A., and Ghazisaedy, M. (2015). The impact of social robotics on L2 learners' anxiety and attitude in english vocabulary acquisition. *International Journal of Social Robotics*, 7(4), pp. 523–535.

Alemi, M., Meghdari, A., and Haeri, N. S. (2017). Young EFL learners' attitude towards rall: An observational study focusing on motivation, anxiety, and interaction. In *Proceedings of the 8th International Conference on Social Robotics (ICSR '17)*, Tsukuba, Japan, pp. 252–261.

Aleven, V. (2010). *Rule-Based Cognitive Modeling for Intelligent Tutoring Systems*, pp. 33–62. Springer Berlin Heidelberg, Berlin, Heidelberg.

Alexander, S., Sarrafzadeh, A., and Hill, S. (2008). Foundation of an affective tutoring system: Learning how human tutors adapt to student emotion. *International journal of Intelligent Systems Technologies and Applications*, 4(3-4), pp. 355–367.

Alexander, S., Sarrafzadeh, A., Hill, S., et al. (2006). Easy with eve: A functional affective tutoring system. In *Proceedings of the 8th Workshop on Motivational and Affective Issues in ITS at the 8th International Conference on ITS (ITS '06)*, Jhongli, Taiwan, pp. 5–12.

Alm, C. O., Roth, D., and Sproat, R. (2005). Emotions from text: Machine learning for text-based emotion prediction. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT and EMNLP '05)*, Vancouver, British Columbia, Canada, pp. 579–586. DOI:10.3115/1220575.1220648.

Altuwairqi, K., Jarraya, S. K., Allinjawi, A., and Hammami, M. (2018). A new emotion–based affective model to detect student's engagement. *Journal of King Saud University - Computer and Information Sciences*. DOI:10.1016/j.jksuci.2018.12.008.

Alves-Oliveira, P., Sequeira, P., Melo, F. S., Castellano, G., and Paiva, A. (2019). Empathic robot for group learning: A field study. *ACM Transactions on Human-Robot Interaction*, 8(1), pp. 3:1–3:34. DOI:10.1145/3300188.

Aminuddin, R., Sharkey, A., and Levita, L. (2016). Interaction with the paro robot may reduce psychophysiological stress responses. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI '16)*, Christchurch, New Zealand, pp. 593–594. DOI:10.1109/HRI.2016.7451872.

Anderson, J. R., Corbett, A. T., Koedinger, K. R., and Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4(2), pp. 167–207. DOI:10.1207/s15327809jls0402_2.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., and Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman, New York City, New York, USA.

Arroyo, I., Beal, C. R., Murray, T., Walles, R., and Woolf, B. P. (2004). Wayang outpost: Intelligent tutoring for high stakes achievement tests. In Lester, J. C., Vicari, R. M., and Paraguaçu, F., editors, *Proceedings of the 7th International Conference on Intelligent Tutoring Systems (ITS '04)*, Alagoas, Brazil, pp. 31–63.

Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., and Christopherson, R. (2009). Emotion sensors go to school. In *Proceedings of the 2009 Conference on Artificial Intelligence in Education (AIED '09)*, volume 200, Brighton, UK, pp. 17–24.

Bandura, A. (1991). Self-regulation of motivation through anticipatory and self-reactive mechanisms. In Dienstbier, R. A., editor, *Current theory and research in motivation, Vol. 38. Nebraska Symposium on Motivation: Perspectives on motivation*, volume 38, Lincoln, Nebraska, USA, pp. 69–164.

Banks, M. R., Willoughby, L. M., and Banks, W. A. (2008). Animal-assisted therapy and loneliness in nursing homes: use of robotic versus living dogs. *Journal of the American Medical Directors Association*, 9(3), pp. 173–177.

Bänziger, T., Grandjean, D., and Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The multimodal emotion recognition test (mert). *Emotion*, 9(5), pp. 691–704.

Barrow, I. M., Holbert, D., and Rastatter, M. P. (2000). Effect of color on developmental picture-vocabulary naming of 4-, 6-, and 8-year-old children. *Speech-Language Pathology*, 9(4), pp. 310–318.

Bartneck, C. and Forlizzi, J. (2004). A design-centred framework for social human-robot interaction. In *Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication (RO-MAN '04)*, Kurashiki, Okayama, Japan, pp. 591–594.

Basawapatna, A. R., Repenning, A., Koh, K. H., and Nickerson, H. (2013). The zones of proximal flow: Guiding students through a space of computational thinking skills and challenges. In *Proceedings of the 9th Annual International ACM Conference on International Computing Education Research (ICER '13)*, San Diego, San California, USA, pp. 67–74. DOI:10.1145/2493394.2493404.

Begum, M., Serna, R. W., and Yanco, H. A. (2016). Are robots ready to deliver autism interventions? A comprehensive review. *International Journal of Social Robotics*, 8(2), pp. 157–181. DOI:10.1007/s12369-016-0346-y.

Belpaeme, T., Baxter, P., de Greeff, J., Kennedy, J., Read, R., Looije, R., Neerincx, M., Baroni, I., and Zelati, M. C. (2013a). Child-robot interaction: Perspectives and challenges. In *Proceedings of the 4th International Conference on Social Robotics (ICSR '13)*, Bristol, UK, pp. 452–459.

Belpaeme, T., Baxter, P., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B., Racioppa, S., Kruijff-Korbayová, I., Athanasopoulos, G., Enescu, V., et al. (2013b). Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, 1(2), pp. 33–53.

Belpaeme, T., Kennedy, J., Baxter, P., Vogt, P., Krahmer, E. E., Kopp, S., Bergmann, K., Leseman, P., Küntay, A. C., Göksun, T., et al. (2015). L2TOR – second language tutoring using social robots. In *Proceedings of the WONDER Workshop at the 7th International Conference on Social Robotics (ICSR '15)*, Paris, France.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018a). Social robots for education: A review. *Science Robotics*, 3(21). DOI:10.1126/scirobotics.aat5954.

Belpaeme, T., Vogt, P., van den Berghe, R., Bergmann, K., Göksun, T., de Haas, M., Kanero, J., Kennedy, J., Küntay, A. C., Oudgenoeg-Paz, O., Papadopoulos, F., Schodde, T., Verhagen, J., Wallbridge, C. D., Willemsen, B., de Wit, J., Geçkin, V., Hoffmann, L., Kopp, S., Krahmer, E., Mamus, E., Montanier, J.-M., Oranç, C., and Pandey, A. K. (2018b). Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics*, 10(3), pp. 325–341. DOI:10.1007/s12369-018-0467-6.

Bennane, A. (2002). An approach of reinforcement learning use in tutoring systems. In Cerri, S. A., Gouardères, G., and Paraguaçu, F., editors, *Proceedings of the 5th International Conference on Intelligent Tutoring Systems (ITS '02)*, San Sebastián, Spain, pp. 993–993.

Bennane, A. (2013). Adaptive educational software by applying reinforcement learning. *Informatics in Education*, 12(1), pp. 13–27.

Benta, K.-I. and Veida, M. (2015). Towards real-life facial expression recognition systems. *Advances in Electrical and Computer Engineering*, 15(2), pp. 93–102.

Bergmann, K., Hoffmann, L., Kopp, S., and Schodde, T. (2017). L2TOR Project Deliverable D5.1: Interaction management for the number domain. http://www.l2tor.eu/effe/wp-content/uploads/2015/12/D5.1-Interaction-management-for-the-number-domain.pdf.

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. McGraw-Hill Book Company, New York City, New York, USA.

Bialystok, E. (2007). Cognitive effects of bilingualism: How linguistic experience leads to cognitive change. *International Journal of Bilingual Education and Bilingualism*, 10(3), pp. 210–223.

Blasco, M. E. (2014). Applying semantic frames to effective vocabulary teaching in the EFL classroom. *Fòrum de Recerca*, 19, pp. 743–752.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6), pp. 4–16.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., and Krathwohl, D. R. (1956). *Taxonomy of educational objectives, Handbook I: Cognitive Domain*. Addison-Wesley Longman Ltd, Boston, MA, USA, 2 edition.

Blumenfeld, P. C., Kempler, T. M., and Krajcik, J. S. (2005). *Motivation and Cognitive Engagement in Learning Environments*, chapter 28, pp. 475–488. Cambridge University Press, Cambridge, UK.

Bohus, D. and Horvitz, E. (2009). Models for multiparty engagement in open-world dialog. In *Proceedings of the 10th Annual SIGDIAL Conference on Discourse and Dialogue (SIGDIAL '09)*, London, UK, pp. 225–234.

Bransford, J., Brown, A., and Cocking, R. E. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. National Academies Press, Washington, DC, USA.

Broadbent, E., Garrett, J., Jepsen, N., Ogilvie, V. L., Ahn, H. S., Robinson, H., Peri, K., Kerse, N., Rouse, P., Pillai, A., et al. (2018). Using robots at home to support patients with chronic obstructive pulmonary disease: pilot randomized controlled trial. *Journal of medical Internet research*, 20(2). DOI:10.2196/jmir.8640.

Broadbent, E., Peri, K., Kerse, N., Jayawardena, C., Kuo, I., Datta, C., and MacDonald, B. (2014). Robots in older people's homes to improve medication adherence and quality of life: A randomised cross-over trial. In *Proceedings of the 6th International Conference on Social Robotics (ICSR '14)*, Sydney, New South Wales, Australia, pp. 64–73.

Bruner, J. (1978). The role of dialogue in language acquisition. *The Child's Conception of Language*, 2(3), pp. 241–256.

Brunskill, E. and Russell, S. J. (2011). Partially observable sequential decision making for problem selection in an intelligent tutoring system. In *Proceedings of the 4th International Conference on Educational Data Mining (EDM '11)*, Eindhoven, the Netherlands, pp. 327–328.

Bull, S., Jackson, T. J., and Lancaster, M. J. (2010). Students' interest in their misconceptions in first-year electrical circuits and mathematics courses. *Electrical Engineering Education*, 47(3), pp. 307–318. DOI:10.7227/IJEEE.47.3.6.

Bull, S. and Kay, J. (2010). Open learner models. In Nkambou, R., Bourdeau, J., and Mizoguchi, R., editors, *Advances in Intelligent Tutoring Systems*, pp. 301–322. Springer Berlin Heidelberg, Berlin, Heidelberg.

Busso, C., Deng, Z., Yildirim, S., Bulut, M., Lee, C. M., Kazemzadeh, A., Lee, S., Neumann, U., and Narayanan, S. (2004). Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th International Conference on Multimodal interfaces (ICMI '04)*, State College, Pennsylvania, USA, pp. 205–211. DOI:10.1145/1027933.1027968.

Cakmak, M. and Lopes, M. (2012). Algorithmic and human teaching of sequential decision tasks. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI '12)*, Toronto, Ontario, Canada, pp. 1536–1542.

Capone, N. C. and McGregor, K. K. (2005). The effect of semantic representation on toddlers' word retrieval. *Journal of Speech, Language, and Hearing Research*, 48(6), pp. 1468–1480.

Caspi, A. and Blau, I. (2008). Social presence in online discussion groups: Testing three conceptions and their relations to perceived learning. *Social Psychology of Education*, 11(3), pp. 323–346.

Castellano, G., Leite, I., Pereira, A., Martinho, C., Paiva, A., and McOwan, P. W. (2012). Detecting engagement in hri: An exploration of social and task-based context. In *Proceedings of the 2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Confernece on Social Computing*, Amsterdam, The Netherlands, pp. 421–428. DOI:10.1109/SocialCom-PASSAT.2012.51.

Castellano, G., Pereira, A., Leite, I., Paiva, A., and McOwan, P. W. (2009). Detecting user engagement with a robot companion using task and social interaction-based features. In *Proceedings of the 6th Workshop on Machine Learning for Multimodal Interactionthe at the 11th International Conference on Multimodal Interfaces (ICMI-MLMI '09)*, Cambridge, UK, pp. 119–126. DOI:10.1145/1647314.1647336.

Castellano, M., Mastronardi, G., di Giuseppe, G., and Dicensi, V. (2007). Neural techniques to improve the formative evaluation procedure in intelligent tutoring systems. In *Proceedings of 2007 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications (CIVEMSA '07)*, Ostuni, Italy, pp. 63–67. DOI:10.1109/CIMSA.2007.4362540.

Cen, H., Koedinger, K., and Junker, B. (2006). Learning factors analysis – a general method for cognitive model evaluation and improvement. In Ikeda, M., Ashley, K. D., and Chan, T.-W., editors, *Proceedings of the 7th International Conference on Intelligent Tutoring Systems (ITS '06)*, Jhongli, Taiwan, pp. 164–175.

Cha, H. J., Kim, Y. S., Park, S. H., Yoon, T. B., Jung, Y. M., and Lee, J.-H. (2006). Learning styles diagnosis based on user interface behaviors for the customization of learning interfaces in an intelligent tutoring system. In Ikeda, M., Ashley, K. D., and Chan, T.-W., editors, *Proceedings of the 7th International Conference on Intelligent Tutoring Systems (ITS '06)*, Jhongli, Taiwan, pp. 513–524.

Chang, W.-L. and Šabanović, S. (2015). Studying socially assistive robots in their organizational context: Studies with paro in a nursing home. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts (HRI '15)*, Portland, Oregon, USA, pp. 227–228. DOI:10.1145/2701973.2702722.

Chen, L., Admoni, H., and Dubrawski, A. (2018). Toward a companion robot fostering perseverance in math-a pilot study. In *Proceedings at the Robots for Learning Workshop at the 13th International Conference of Human-Robot Interaction (R4L '18)*, Chicago, Illinois, USA.

Chi, M., Jordan, P., Vanlehn, K., and Litman, D. (2009). To elicit or to tell: Does it matter? In *Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED '09)*, Amsterdam, The Netherlands, pp. 197–204.

Chi, M., Koedinger, K. R., Gordon, G. J., Jordon, P., and VanLahn, K. (2011). Instructional factors analysis: A cognitive model for multiple instructional interventions. In *Proceedings of the 4th International Conference on Educational Data Mining (EDM '04)*, Einhoven, The Netherlands, pp. 61–70.

Christ, T. and Wang, X. C. (2012). Supporting preschoolers' vocabulary learning: Using a decision-making model to select appropriate words and methods. *Young Children*, 67(2), pp. 74–80.

Clabaugh, C., Ragusa, G., Sha, F., and Matarić, M. (2015). Designing a socially assistive robot for personalized number concepts learning in preschool children. In *Proceedings of the 5th Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EPIROB '15)*, Providence, Rhode Island, USA, pp. 314–319. DOI:10.1109/DEVLRN.2015.7346164.

Clancey, W. J. (1984). Use of mycin's rules for tutoring. *Rule-Based Expert Systems*.

Clement, B., Roy, D., Oudeyer, P., and Lopes, M. (2015). Multi-armed bandits for intelligent tutoring systems. *Educational Data Mining*, 7(2), pp. 20–48.

Cohen, P. A., Kulik, J. A., and Kulik, C.-L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19(2), pp. 237–248. DOI:10.3102/00028312019002237.

Conati, C. (2002). Probabilistic assessment of user's emotions in educational games. *Applied Artificial Intelligence*, 16(7-8), pp. 555–575. DOI:10.1080/08839510290030390.

Conati, C. (2009). Intelligent tutoring systems: new challenges and directions. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI '09)*, Pasadena, California, USA, pp. 2–7.

Conati, C., Gertner, A., and VanLehn, K. (2002). Using bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction*, 12(4), pp. 371–417. DOI:10.1023/A:1021258506583.

Conati, C. and Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, 19(3), pp. 267–303. DOI:10.1007/s11257-009-9062-8.

Corbalan, G., Kester, L., and van Merriënboer, J. J. (2009). Dynamic task selection: Effects of feedback and learner control on efficiency and motivation. *Learning and Instruction*, 19(6), pp. 455–465. DOI:10.1016/j.learninstruc.2008.07.002.

Corbett, A. T. and Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), pp. 253–278. DOI:10.1007/BF01099821.

Core, M. G. and Allen, J. (1997). Coding dialogs with the damsl annotation scheme. In *Proceedings of the AAAI fall symposium on communicative action in humans and machines*, volume 56, Cambridge, UK.

Couture-Beil, A., Vaughan, R. T., and Mori, G. (2010). Selecting and commanding individual robots in a multi-robot system. In *Proceedings of the 7th Canadian Conference on Computer and Robot Vision (CRV '10)*, Ottawa, Ontario, Canada, pp. 159–166. DOI:10.1109/CRV.2010.28.

Craig, S., Graesser, A., Sullins, J., and Gholson, B. (2004). Affect and learning: An exploratory look into the role of affect in learning with autotutor. *Educational Media*, 29(3), pp. 241–250. DOI:10.1080/1358165042000283101.

Crowley, R. S. and Medvedeva, O. (2006). An intelligent tutoring system for visual classification problem solving. *Artificial Intelligence in Medicine*, 36(1), pp. 85–117. DOI:10.1016/j.artmed.2005.01.005.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimal Experience*. Harper and Row, New York City, New York, USA.

Csikszentmihalyi, M. (1996). *Flow and the psychology of discovery and invention*, volume 56. Harper Collins, New York City, New York, USA.

Csikszentmihalyi, M. (2014). *Flow*. Springer Netherlands, Dordrecht, The Netherlands.

Culatta, R. (2011). Zone of proximal development. `http://www.innovativelearning.com/educational_psychology/development/zone-of-proximal-development.html`. Last checked on June 11, 2019.

d Baker, R. S., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., Salvi, A., Aleven, V., Kusbit, G. W., Ocumpaugh, J., and Rossi, L. (2012). Towards sensor-free affect detection in cognitive tutor algebra. In *Proceedings of the 5th International Conference on Educational Data Mining (EDM '12)*, Chania, Greece, pp. 126–133.

Dahlbäck, N., Jönsson, A., and Ahrenberg, L. (1993). Wizard of oz studies: Why and how. In *Proceedings of the 1st International Conference on Intelligent User Interfaces (IUI '93)*, Orlando, Florida, USA, pp. 193–200. DOI:10.1145/169891.169968.

Dale, P. S., Harlaar, N., and Plomin, R. (2012). Nature and nurture in school-based second language achievement. *Language Learning*, 62(s2), pp. 28–48. DOI:10.1111/j.1467-9922.2012.00705.x.

Dautenhahn, K. and Werry, I. (2004). Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, 12(1), pp. 1–35.

Dave, R. H. (1970). Psychomotor levels. In Armstrong, R., editor, *Developing and Writing Behavioral Objectives*, pp. 20–21. Educational Innovators Press, Tucson, Arizona, USA.

David Cornish, M., Dukette, D., et al. (2009). *The essential 20: Twenty components of an excellent health care team*. Dorrance Publishing, Pittsburgh, Pennsylvania, USA.

Davidson, J. E., Sternberg, R. J., and Sternberg, R. J. (2003). *The psychology of problem solving*. Cambridge University Press, Cambridge, UK.

Davidson, R. J. and Cacioppo, J. T. (1992). New developments in the scientific study of emotion: An introduction to the special section. *Psychological Science*, 3(1), pp. 21–22. DOI:10.1111/j.1467-9280.1992.tb00250.x.

Davies, N. (2016). Can robots handle your healthcare? *Engineering Technology*, 11(9), pp. 58–61. DOI:10.1049/et.2016.0907.

de Baker, R. S. J., Corbett, A. T., and Aleven, V. (2008). Improving contextual models of guessing and slipping with a truncated training set. In *Proceedings of the 1st International Conference on Educational Data Mining (EDM '08)*, Montreal, Quebec, Canada, pp. 67–76.

de Nooijer, J. A., van Gog, T., Paas, F., and Zwaan, R. A. (2013). Effects of imitating gestures during encoding or during retrieval of novel verbs on children's test performance. *Acta Psychologica*, 144(1), pp. 173–179. DOI:10.1016/j.actpsy.2013.05.013.

de Vicente, M. (2018). Cognifit - health, brain & neuroscience. `https://blog.cognifit.com/cognitive-learning-an-education-guide-to-types-of-learning/`. Last checked on June 11, 2019.

de Wit, J., Schodde, T., Willemsen, B., Bergmann, K., de Haas, M., Kopp, S., Krahmer, E., and Vogt, P. (2017). Exploring the effect of gestures and adaptive tutoring on children's comprehension of L2 vocabularies. In *Proceedings of Robots for Learning Workshop at 12th International Conference of Human-Robot Interaction (R4L '17)*, Vienna, Austria.

de Wit, J., Schodde, T., Willemsen, B., Bergmann, K., de Haas, M., Kopp, S., Krahmer, E., and Vogt, P. (2018). The effect of a robot's gestures and adaptive tutoring on children's acquisition of second language vocabularies. In *Proceedings of the 13th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*, Chicago, Illinois, USA, pp. 50–58. DOI:10.1145/3171221.3171277.

Dede, C. (1986). A review and synthesis of recent research in intelligent computer-assisted instruction. *International Journal of Man-Machine Studies*, 24(4), pp. 329–353.

del Moral, S., Pardo, D., and Angulo, C. (2009). Social robot paradigms: An overview. In *Proceedings of the 10th International Work-Conference on Artificial Neural Networks*, Salamanca, Spain, pp. 773–780.

Derry, S., Hawkes, L., and Ziegler, U. (1988). A plan-based opportunistic architecture for intelligent tutoring. In *Proceedings of the 1th International Conference in Intelligent Tutoring Systems (ITS '88)*, Montreal, Quebec, Canada, pp. 116–123.

Devillers, L. and Vidrascu, L. (2006). Real-life emotions detection with lexical and paralinguistic cues on human-human call center dialogs. In *Proceedings of the 9th International Conference on Spoken Language Processing (ICSLP '06)*, Pittsburgh, Pennsylvania, USA, pp. 801–804.

D'Mello, S. and Graesser, A. (2009). Automatic detection of learner's affect from gross body language. *Applied Artificial Intelligence*, 23(2), pp. 123–150.

D'Mello, S. and Graesser, A. (2012). Autotutor and affective autotutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems - Special issue on Highlights of the Decade in Interactive Intelligent Systems*, 2(4), pp. 23:1 –23:39.

D'Mello, S. K., Craig, S. D., Witherspoon, A., McDaniel, B., and Graesser, A. (2008). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction*, 18(1), pp. 45–80. DOI:10.1007/s11257-007-9037-6.

Dresel, M. and Haugwitz, M. (2008). A computer-based approach to fostering motivation and self-regulated learning. *The Journal of Experimental Education*, 77(1), pp. 3–20. DOI:10.3200/JEXE.77.1.3-20.

El Haddioui, I. and Khaldi, M. (2012). Learner behavior analysis through eye tracking. *International Journal of Computer Science Research and Application*, 2, pp. 11–18.

Ellis, D. and Zimmerman, B. J. (2001). *Enhancing Self-Monitoring during Self-Regulated Learning of Speech*, pp. 205–228. Springer Netherlands, Dordrecht, The Netherlands.

Engeser, S. and Rheinberg, F. (2008). Flow, performance and moderators of challenge-skill balance. *Motivation and Emotion*, 32(3), pp. 158–172.

Esposito, A. (2009). *Affect in Multimodal Information*, pp. 203–226. Springer London, London, UK.

Esser, H. (2006). *Migration, language and integration*. WZB, Berlin, Germany.

Fasola, J. and Matarić, M. J. (2013). A socially assistive robot exercise coach for the elderly. *Journal of Human-Robot Interaction*, 2(2), pp. 3–32.

Feil-Seifer, D. and Mataric, M. J. (2005). Defining socially assistive robotics. In *Proceedings of the 9th International Conference on Rehabilitation Robotics (ICORR '05)*, Chicago, Illinois, USA, pp. 465–468.

Ferguson, K., Arroyo, I., Mahadevan, S., Woolf, B., and Barto, A. (2006). Improving intelligent tutoring systems: Using expectation maximization to learn student skill levels. In Ikeda, M., Ashley, K. D., and Chan, T.-W., editors, *Proceedings of the 9th International Conference on Intelligent Tutoring Systems (ITS '06)*, Jhongli, Taiwan, pp. 453–462.

Folsom-Kovarik, J. T., Sukthankar, G., and Schatz, S. (2013). Tractable pomdp representations for intelligent tutoring systems. *ACM Transactions on Intelligent Systems and Technology*, 4(2), pp. 29:1–29:22. DOI:10.1145/2438653.2438664.

Folsom-Kovarik, J. T., Sukthankar, G., Schatz, S. L., and Nicholson, D. M. (2010). Scalable pomdps for diagnosis and planning in intelligent tutoring systems. In *Proceedings of the 2010 AAAI Fall Symposium: Proactive Assistant Agents*, Arlington, Virginia, USA, pp. 2–7.

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems: Concepts, Design, and Applications*, 42(3-4), pp. 143–166.

Foster, M. E., Gaschler, A., and Giuliani, M. (2017). Automatically classifying user engagement for dynamic multi-party human–robot interaction. *International Journal of Social Robotics*, 9(5), pp. 659–674. DOI:10.1007/s12369-017-0414-y.

Freedman, R., Ali, S. S., and McRoy, S. (2000). Links: What is an intelligent tutoring system? *Intelligence*, 11(3), pp. 15–16. DOI:10.1145/350752.350756.

Fridin, M. (2014a). Kindergarten social assistive robot: First meeting and ethical issues. *Computers in Human Behavior*, 30, pp. 262–272. DOI:10.1016/j.chb.2013.09.005.

Fridin, M. (2014b). Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. *Computers & Education*, 70, pp. 53–64. DOI:10.1016/j.compedu.2013.07.043.

Gairns, R., Redman, S., et al. (1986). *Working with words: A guide to teaching and learning vocabulary*. Cambridge University Press Cambridge, Cambridge, UK.

Ghadirli, H. M. and Rastgarpour, M. (2013). A web-based adaptive and intelligent tutor by expert systems. In *Advances in Computing and Information Technology - Proceedings of the 2nd International Conference on Advances in Computing and Information Technology*, Chennai, India, pp. 87–95.

Ghasemi, B. and Hashemi, M. (2011). Foreign language learning during childhood. *Procedia - Social and Behavioral Sciences*, 28, pp. 872–876. DOI:10.1016/j.sbspro.2011.11.160.

Gibbons, P. (2002). *Scaffolding language, scaffolding learning*. Portsmouth, NH: Heinemann, Berkeley, California, USA.

Glas, N. and Pelachaud, C. (2014). Politeness versus perceived engagement: an experimental study. In *Proceedings of the 11th Workshop on Natural Language Processing and Cognitive Science (NLPCS '14)*, Venice, Italy, pp. 135–144.

Glas, N. and Pelachaud, C. (2015). Definitions of engagement in human-agent interaction. In *Proceedings of the 6th International Conference on Affective Computing and Intelligent Interaction (ACII '15)*, Xi'an, China, pp. 944–949. DOI:10.1109/ACII.2015.7344688.

Goffman, E. (1966). *Behavior in Public Places: Notes on the Social Organization of Gatherings*. Simon and Schuster, New York City, New York, USA.

Gong, Y., Beck, J. E., and Heffernan, N. T. (2010). Comparing knowledge tracing and performance factor analysis by using multiple model fitting procedures. In Aleven, V., Kay, J., and Mostow, J., editors, *Proceedings of the 9th International Conference on Intelligent Tutoring Systems (ITS '10)*, Pittsburgh, Pennsylvania, USA, pp. 35–44.

Gonzalez-Brenes, J., Huang, Y., and Brusilovsky, P. (2014). General features in knowledge tracing to model multiple subskills, temporal item response theory, and expert knowledge. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM '14)*, London, UK, pp. 84–91.

Gordon, G. and Breazeal, C. (2015). Bayesian active learning-based robot tutor for children's word-reading skills. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI '15)*, Austin, Texas, USA, pp. 1343–1349.

Gordon, G., Spaulding, S., Westlund, J. K., Lee, J. J., Plummer, L., Martinez, M., Das, M., and Breazeal, C. (2016). Affective personalization of a social robot tutor for children's second language skills. In *Proceedings of 30th AAAI Conference on Artificial Intelligence (AAAI '16)*, Phoenix, Arizona, USA, pp. 3951–3957.

Gorham, J. (1988). The relationship between verbal teacher immediacy behaviors and student learning. *Communication Education*, 37(1), pp. 40–53. DOI:10.1080/03634528809378702.

Graesser, A. C., Conley, M. W., and Olney, A. (2012). Intelligent tutoring systems. In *APA educational psychology handbook, Vol 3: Application to learning and teaching*, pp. 451–473. American Psychological Association, Washington, DC, USA.

Graesser, A. C., Wiemer-Hastings, P., Wiemer-Hastings, K., Harter, D., Group, T. R. G. T. R., and Person, N. (2000). Using latent semantic analysis to evaluate the contributions of students in autotutor. *Interactive Learning Environments*, 8(2), pp. 129–147. DOI:10.1076/1049-4820(200008)8:2;1-B;FT129.

Greczek, J., Kaszubski, E., Atrash, A., and Matarić, M. (2014). Graded cueing feedback in robot-mediated imitation practice for children with autism spectrum disorders. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '14)*, Edinburgh, UK, pp. 561–566. DOI:10.1109/ROMAN.2014.6926312.

Greene, B. A. and Miller, R. B. (1996). Influences on achievement: Goals, perceived ability, and cognitive engagement. *Contemporary Educational Psychology*, 21(2), pp. 181–192. DOI:10.1006/ceps.1996.0015.

Haas, M. d., Baxter, P., de Jong, C., Krahmer, E., and Vogt, P. (2017). Exploring different types of feedback in preschooler and robot interaction. In *Proceedings of the Companion Workshop at the 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*, Vienna, Austria, pp. 127–128. DOI:10.1145/3029798.3038433.

Habgood, M. J. and Ainsworth, S. E. (2011). Motivating children to learn effectively: Exploring the value of intrinsic integration in educational games. *The Journal of the Learning Sciences*, 20(2), pp. 169–206.

Hall, L., Woods, S., Aylett, R., Newall, L., and Paiva, A. (2005). Achieving empathic engagement through affective interaction with synthetic characters. In *Proceedings of the 1st international conference on Affective Computing and Intelligent Interaction (ACII '05)*, Beijing, China, pp. 731–738.

Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., and Edwards, T. (2016). Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*, 54, pp. 170–179. DOI:10.1016/j.chb.2015.07.045.

Han, J., Jo, M., Park, S., and Kim, S. (2005). The educational use of home robots for children. In *Proceedings of the 14th International Workshop on Robot and Human Interactive Communication (RO-MAN '05).*, Nashville, Tennessee, USA, pp. 378–383. DOI:10.1109/ROMAN.2005.1513808.

Han, J.-H., Jo, M.-H., Jones, V., and Jo, J.-H. (2008). Comparative study on the educational use of home robots for children. *Journal of Information Processing Systems*, 4(4), pp. 159–168.

Harrow, A. (1972). *A Taxonomy of Psychomotor Domain: A Guide for Developing Behavioral Objectives*. David McKay, New York City, New York, USA.

Hartmann, H. (2002). *Human learning and instruction*. City University of New York, New York City, New York, USA.

Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), pp. 81–112.

Hatzilygeroudis, I. and Prentzas, J. (2004). Using a hybrid rule-based approach in developing an intelligent tutoring system with knowledge acquisition and update capabilities. *Expert Systems with Applications*, 26(4), pp. 477–492. DOI:10.1016/j.eswa.2003.10.007.

Henkemans, O. A. B., Bierman, B. P., Janssen, J., Neerincx, M. A., Looije, R., van der Bosch, H., and van der Giessen, J. A. (2013). Using a robot to personalise health education for children with diabetes type 1: A pilot study. *Patient Education and Counseling*, 92(2), pp. 174–181. DOI:10.1016/j.pec.2013.04.012.

Herberg, J. S., Feller, S., Yengin, I., and Saerbeck, M. (2015). Robot watchfulness hinders learning performance. In *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '15)*, Kobe, Japan, pp. 153–160. DOI:10.1109/ROMAN.2015.7333620.

Hockema, S. A. and Smith, L. B. (2009). Learning your language, outside-in and inside-out. *Linguistics*, 47(2), pp. 453–479.

Hoff, E. (2013). Interpreting the early language trajectories of children from low-ses and language minority homes: implications for closing achievement gaps. *Developmental psychology*, 49(1), pp. 4–14. DOI:10.1037/a0027238.

Hogan, K. and Pressley, M. (1997). *Scaffolding Student Learning: Instructional Approaches and Issues*. Advances in teaching and learning. Brookline Books, Northampton, Massachusetts, USA.

Hooper, S. R. and Umansky, W. (2009). *Young children with special needs*. Pearson Merrill Prentice Hall, Upper Saddle River, New Jersey, USA.

Horst, J. S. and Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, 13(2), pp. 128–157.

Hsiao, H.-S., Chang, C.-S., Lin, C.-Y., and Hsu, H.-L. (2015). "irobiq": the influence of bidirectional interaction on kindergarteners' reading motivation, literacy, and behavior. *Interactive Learning Environments*, 23(3), pp. 269–292.

Huijnen, C. A. G. J., Lexis, M. A. S., and de Witte, L. P. (2016). Matching robot kaspar to autism spectrum disorder (asd) therapy and educational goals. *International Journal of Social Robotics*, 8(4), pp. 445–455. DOI:10.1007/s12369-016-0369-4.

Huskens, B., Verschuur, R., Gillesen, J., Didden, R., and Barakova, E. (2013). Promoting question-asking in school-aged children with autism spectrum disorders: Effectiveness of a robot intervention compared to a human-trainer intervention. *Developmental Neurorehabilitation*, 16(5), pp. 345–356.

Hyun, E.-j., Kim, S.-y., Jang, S., and Park, S. (2008). Comparative study of effects of language instruction program using intelligence robot and multimedia on linguistic ability of young children. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '08)*, Munich, Germany, pp. 187–192.

Immordino-Yang, M. H. and Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, brain, and education*, 1(1), pp. 3–10.

Jacq, A., Lemaignan, S., Garcia, F., Dillenbourg, P., and Paiva, A. (2016). Building successful long child-robot interactions in a learning context. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI '16)*, Christchurch, New Zealand, pp. 239–246. DOI:10.1109/HRI.2016.7451758.

Jampert, K., Leuckenfeld, K., Zehnbauer, A., and Best, P. (2006). *Sprachliche Förderung der Kita : Wie viel Sprache steckt in Musik, Bewegung, Naturwissenschaften und Medien?* verlag das netz, Kiliansroda, Germany.

Janssen, J. B., van der Wal, C. C., Neerincx, M. A., and Looije, R. (2011). Motivating children to learn arithmetic with an adaptive robot game. In *Proceedings of the 2nd International Conference on Social Robotics (ICSR '11)*, Amsterdam, The Netherlands, pp. 153–162.

Jenkins, S. and Draper, H. (2014). Robots and the division of healthcare responsibilities in the homes of older people. In *Proceedings of the 6th International Conference on Social Robotics (ICSR '14)*, Sydney, NSW, Australia, pp. 176–185.

Johansson, E. (2004). Learning encounters in preschool: Interaction between atmosphere, view of children and of learning. *International journal of early childhood*, 36(2), pp. 9–26.

Johnson, D. O., Cuijpers, R. H., and van der Pol, D. (2013). Imitating human emotions with artificial facial expressions. *International Journal of Social Robotics*, 5(4), pp. 503–513. DOI:10.1007/s12369-013-0211-1.

Jones, A., Bull, S., and Castellano, G. (2017). "I Know That Now, I'm Going to Learn This Next" Promoting Self-regulated Learning with a Robotic Tutor. *International Journal of Social Robotics*, pp. 1–16. DOI:10.1007/s12369-017-0430-y.

Jones, A. and Castellano, G. (2018). Adaptive robotic tutors that support self-regulated learning: A longer-term investigation with primary school children. *International Journal of Social Robotics*, 10(3), pp. 357–370. DOI:10.1007/s12369-017-0458-z.

Jones, A., Castellano, G., and Bull, S. (2014). Investigating the effect of a robotic tutor on learner perception of skill based feedback. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 8755(1), pp. 186–195. DOI:10.1007/978-3-319-11973-1_19.

Jones, A., Küster, D., Basedow, C. A., Alves-Oliveira, P., Serholt, S., Hastie, H., Corrigan, L. J., Barendregt, W., Kappas, A., Paiva, A., and Castellano, G. (2015). Empathic robotic tutors for personalised learning: A multidisciplinary approach. In *Proceedings of the 6th International Conference on Social Robotics (ICSR '15)*, Paris, France, pp. 285–295.

Kahn, J. H., Tobin, R. M., Massey, A. E., and Anderson, J. A. (2007). Measuring emotional expression with the linguistic inquiry and word count. *The American Journal of Psychology*, 120(2), pp. 263–286.

Kanda, T., Hirano, T., Eaton, D., and Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*, 19(1), pp. 61–84. DOI:10.1207/s15327051hci1901&2_4.

Kapoor, A. and Picard, R. W. (2005). Multimodal affect recognition in learning environments. In *Proceedings of the 13th Annual ACM international conference on Multimedia (MULTIMEDIA '05)*, Hilton, Singapore, pp. 677–682. DOI:10.1145/1101149.1101300.

Käser, T., Klingler, S., Schwing, A. G., and Gross, M. (2014a). Beyond knowledge tracing: Modeling skill topologies with bayesian networks. In Trausan-Matu, S., Boyer, K. E., Crosby, M., and Panourgia, K., editors, *Proceedings of the 12th International Conference on Intelligent Tutoring Systems (ITS '14)*, Honolulu, Hawaii, USA, pp. 188–198.

Käser, T., Koedinger, K. R., and Gross, M. H. (2014b). Different parameters - same prediction: An analysis of learning curves. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM '14)*, London, UK, pp. 52–59.

Kennedy, J., Baxter, P., and Belpaeme, T. (2015). The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*, Portland, Oregon, USA, pp. 67–74. DOI:10.1145/2696454.2696457.

Kennedy, J., Baxter, P., and Belpaeme, T. (2017a). Nonverbal immediacy as a characterisation of social behaviour for human–robot interaction. *International Journal of Social Robotics*, 9(1), pp. 109–128. DOI:10.1007/s12369-016-0378-3.

Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2016). Social robot tutoring for child second language learning. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI '16)*, Christchurch, New Zealand, pp. 231–238. DOI:10.1109/HRI.2016.7451757.

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., Senft, E., and Belpaeme, T. (2017b). Child speech recognition in human-robot interaction: evaluations and recommendations. In *Proceedings of the 12th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*, Vienna, Austria, pp. 82–90.

Kersten, A. W. and Smith, L. B. (2002). Attention to novel objects during verb learning. *Child development*, 73(1), pp. 93–109.

Khajah, M., Lindsey, R. V., and Mozer, M. C. (2016). How deep is knowledge tracing? In *Proceedings of the 9th International Conference on Educational Data Mining (EDM '16)*, Raleigh, North Carolina, USA.

Khajah, M., Wing, R., Lindsey, R. V., and Mozer, M. C. (2014a). Integrating latent-factor and knowledge-tracing models to predict individual differences in learning. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM '14)*, London, UK, pp. 99–106.

Khajah, M. M., Huang, Y., González-Brenes, J. P., Mozer, M. C., and Brusilovsky, P. (2014b). Integrating knowledge tracing and item response theory: A tale of two frameworks. In *Proceedings of the 4th International Workshop on Personalization Approaches in Learning Environments (PALE '14)*, volume 1181, Aalborg, Denmark, pp. 7–15.

Khoshsima, H., Saed, A., and Yazdani, A. (2015). Instructional games and vocabulary enhancement: Case of iranian pre-intermediate EFL learners. *International Journal of Language and Linguistics*, 3(6), pp. 328–332.

Kidd, C. D. and Breazeal, C. (2004). Effect of a robot on user perceptions. In *Proceedings of the 17th IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '04)*, Sendai, Japan, pp. 3559–3564. DOI:10.1109/IROS.2004.1389967.

Kim, J. C., Azzi, P., Jeon, M., Howard, A. M., and Park, C. H. (2017). Audio-based emotion estimation for interactive robotic therapy for children with autism spectrum disorder. In *Proceedings of the 14th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI '17)*, Jeju, South Korea, pp. 39–44. DOI:10.1109/URAI.2017.7992881.

Kim, S. K. (2007). *Organisation des bilingualen Gehirns: Eine EEG-Studie zum Einfluss von Erberbsalter und Sprachniveau*. PhD thesis, Bielefeld University, Beilefeld, Germany.

Koedinger, K. R. and Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review*, 19(3), pp. 239–264. DOI:10.1007/s10648-007-9049-0.

Kose-Bagci, H., Ferrari, E., Dautenhahn, K., Syrdal, D. S., and Nehaniv, C. L. (2009). Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot. *Advanced Robotics*, 23(14), pp. 1951–1996. DOI:10.1163/016918609X12518783330360.

Kozima, H., Michalowski, M. P., and Nakagawa, C. (2009). Keepon. *International Journal of Social Robotics*, 1(1), pp. 3–18. DOI:10.1007/s12369-008-0009-8.

Kozima, H., Nakagawa, C., and Yasuda, Y. (2007). Children–robot interaction: a pilot study in autism therapy. In von Hofsten, C. and Rosander, K., editors, *From Action to Cognition*, volume 164, pp. 385–400. Elsevier, Amsterdam, The Netherlands.

Krashen, S. D. (1981). *Second Language Acquisition and Second Language learning*. Pergamon Press Inc., Oxford, UK.

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory Into Practice*, 41(4), pp. 212–218. DOI:10.1207/s15430421tip4104_2.

Kratwohl, D. R., Bloom, B. S., and Masia, B. B. (1964). *Taxonomy of Educational Objectives, the classification of educational goals–Handbook II: Affective Domain*. McKay, New York City, New York, USA.

Kuh, G. D. (2003). What we're learning about student engagement from nsse: Benchmarks for effective educational practices. *Change: The Magazine of Higher Learning*, 35(2), pp. 24–32. DOI:10.1080/00091380309604090.

Kuhl, P. K. (2004). Early language acquisition: cracking the speech code. *Nature Reviews Neuroscience*, 5(11), pp. 831–843.

Le Maitre, J. and Chetouani, M. (2013). Self-talk discrimination in human–robot interaction situations for supporting social awareness. *International Journal of Social Robotics*, 5(2), pp. 277–289. DOI:10.1007/s12369-013-0179-x.

Lee, C.-S. (2007). Diagnostic, predictive and compositional modeling with data mining in integrated learning environments. *Computers and Education*, 49(3), pp. 562–580. DOI:10.1016/j.compedu.2005.10.010.

Lee, J. I. and Brunskill, E. (2012). The impact on individualizing student models on necessary practice opportunities. In *Proceedings of the 5th International Conference on Educational Data Mining (EDM '12)*, Chania, Greece, pp. 118–125.

Lee, S., Noh, H., Lee, J., Lee, K., Lee, G. g., Sagong, S., and Kim, M. (2011). On the effectiveness of robot-assisted language learning. *ReCALL*, 23(1), pp. 25–58. DOI:10.1017/S0958344010000273.

Lehman, B., D'Mello, S., and Person, N. (2010). The intricate dance between cognition and emotion during expert tutoring. In Aleven, V., Kay, J., and Mostow, J., editors, *Proceedings of the 10th International Conference on Intelligent Tutoring Systems (ITS '10)*, Pittsburgh, Pennsylvania, USA, pp. 1–10.

Lehmann, H., Syrdal, D. S., Dautenhahn, K., Gelderblom, G., Bedaf, S., and Amirabdollahian, F. (2013). What should a robot do for you?-evaluating the needs of the elderly in the uk. In *Proceedings of the 6th International Conference on Advances in Computer–Human Interactions (ACHI '13)*, Nice, France, pp. 83–88.

Leite, I. (2015). *Long-term Interactions with Empathic Social Robots*. PhD thesis, Univeridade Técnica de Lisboa, Lissabon, Portugal.

Leite, I., Martinho, C., and Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics*, 5(2), pp. 291–308. DOI:10.1007/s12369-013-0178-y.

Leitner, S. (1972). *So lernt man Lernen: Der Weg zum Erfolg [Learning to learn: The road to success]*. Herder, Freiburg, Germany.

Lemaignan, S., Edmunds, C. E. R., Senft, E., and Belpaeme, T. (2018). The PInSoRo dataset: Supporting the data-driven study of child-child and child-robot social dynamics. *PLOS ONE*, 13(10), pp. 1–19. DOI:10.1371/journal.pone.0205999.

Lemaignan, S., Fink, J., Mondada, F., and Dillenbourg, P. (2015). You're doing it wrong! studying unexpected behaviors in child-robot interaction. In Tapus, A., André, E., Martin, J.-C., Ferland, F., and Ammi, M., editors, *Proceedings of the 6th International Conference on Social Robotics (ICSR '15)*, Paris, France, pp. 390–400. DOI:10.1007/978-3-319-25554-5_39.

Lemaignan, S., Garcia, F., Jacq, A., and Dillenbourg, P. (2016a). From real-time attention assessment to "with-me-ness" in human-robot interaction. In *Proceedings of the 11th ACM/IEEE Annual International Conference on Human-Robot Interaction (HRI '16)*, Christchurch, New Zealand, pp. 157–164. DOI:10.1109/HRI.2016.7451747.

Lemaignan, S., Jacq, A., Hood, D., Garcia, F., Paiva, A., and Dillenbourg, P. (2016b). Learning by teaching a robot: The case of handwriting. *IEEE Robotics Automation Magazine*, 23(2), pp. 56–66. DOI:10.1109/MRA.2016.2546700.

Lenneberg, E. H. (1967). The biological foundations of language. *Hospital Practice*, 2(12), pp. 59–67.

Leung, C. B. (1992). Effects of word-related variables on vocabulary growth repeated read-aloud events. *Literacy Research, Theory, and Practice: Views from many perspectives*, pp. 491–498.

Levy, R. I. (1983). Introduction: Self and emotion. *Ethos*, 11(3), pp. 128–134. DOI:10.1525/eth.1983.11.3.02a00020.

Leyzberg, D., Ramachandran, A., and Scassellati, B. (2018). The effect of personalization in longer-term robot tutoring. *ACM Transactions on Human-Robot Interaction*, 7(3), pp. 1–19. DOI:10.1145/3283453.

Leyzberg, D., Spaulding, S., and Scassellati, B. (2014). Personalizing robot tutors to individuals' learning differences. In *Proceedings of the 9th ACM/IEEE International Conference on Human-robot Interaction (HRI '14)*, Bielefeld, Germany, pp. 423–430. DOI:10.1145/2559636.2559671.

Leyzberg, D., Spaulding, S., Toneva, M., and Scassellati, B. (2012). The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, Sapporo, Japan, pp. 1882–1887.

Libin, A. and Cohen-Mansfield, J. (2004). Therapeutic robocat for nursing home residents with dementia: preliminary inquiry. *American Journal of Alzheimer's Disease & Other Dementias®*, 19(2), pp. 111–116.

Lin, X., Hmelo, C., Kinzer, C. K., and Secules, T. J. (1999). Designing technology to support reflection. *Educational Technology Research and Development*, 47(3), pp. 43–62. DOI:10.1007/BF02299633.

Linnenbrink, E. A. and Pintrich, P. R. (2003). The role of self-efficacy beliefs in student engagement and learning in the classroom. *Reading & Writing Quarterly*, 19(2), pp. 119–137. DOI:10.1080/10573560308223.

Lo, A. C., Guarino, P. D., Richards, L. G., Haselkorn, J. K., Wittenberg, G. F., Federman, D. G., Ringer, R. J., Wagner, T. H., Krebs, H. I., Volpe, B. T., Bever, C. T., Bravata, D. M., Duncan, P. W., Corn, B. H., Maffucci, A. D., Nadeau, S. E., Conroy, S. S., Powell, J. M., Huang, G. D., and Peduzzi, P. (2010). Robot-assisted therapy for long-term upper-limb impairment after stroke. *New England Journal of Medicine*, 362(19), pp. 1772–1783. DOI:10.1056/NEJMoa0911341.

Long, Y. and Aleven, V. (2013). Supporting students' self-regulated learning with an open learner model in a linear equation tutor. In *Proceedings of the 2nd International Conference on Artificial Intelligence in Education*, Memphis, Tennessee, USA, pp. 219–228.

Luckin, R. and du Boulay, B. (1999). Ecolab : The development and evaluation of a vygotskian design framework. *Artificial Intelligence in Education*, 10, pp. 198–220.

Lüke, C. and Ritterfeld, U. (2014). The influence of iconic and arbitrary gestures on novel word learning in children with and without SLI. *Gesture*, 14(2), pp. 204–225.

Lyons, J. B. (2013). Being transparent about transparency: A model for human-robot interaction. In Sofge, D., Kruijff, G. J., and Lawless, W. F., editors, *Trust and autonomous systems: Papers from the AAAI Spring Symposium*, Stanford, California, USA, pp. 48–53.

Lyons, J. B., Sadler, G. G., Koltai, K., Battiste, H., Ho, N. T., Hoffmann, L. C., Smith, D., Johnson, W., and Shively, R. (2017). Shaping trust through transparent design: Theoretical and experimental guidelines. In Savage-Knepshield, P. and Chen, J., editors, *Advances in Human Factors in Robots and Unmanned Systems*, volume 499, pp. 127–136. Springer International Publishing, Basel, Switzerland.

Macedonia, M., Müller, K., and Friederici, A. D. (2010). Neural correlates of high performance in foreign language vocabulary learning. *Mind, Brain, and Education*, 4(3), pp. 125–134.

Malpani, A., Ravindran, B., and Murthy, H. (2011). Personalized intelligent tutoring system using reinforcement learning. In *Proceedings of the 24th International Florida Artificial Intelligence Research Society Conference (FLAIRS '11)*, Palm Beach, Florida, USA.

Marilyn, D. (1994). The effect of sign language on hearing children's language development. *Communication Education*, 43(4), pp. 291–298. DOI:10.1080/03634529409378987.

Matarić, M. (2014). Socially assistive robotics: Human-robot interaction methods for creating robots that care. In *Proceedings of the 9th ACM/IEEE International Conference on Human-robot Interaction (HRI '14)*, Bielefeld, Germany, pp. 333–333. DOI:10.1145/2559636.2560043.

Matarić, M. J., Eriksson, J., Feil-Seifer, D. J., and Winstein, C. J. (2007). Socially assistive robotics for post-stroke rehabilitation. *Journal of NeuroEngineering and Rehabilitation*, 4(1), pp. 5. DOI:10.1186/1743-0003-4-5.

Mavilidi, M.-F., Okely, A. D., Chandler, P., Cliff, D. P., and Paas, F. (2015). Effects of integrated physical exercises and gestures on preschool children's foreign language vocabulary learning. *Educational Psychology Review*, 27(3), pp. 413–426. DOI:10.1007/s10648-015-9337-z.

Mayer, K. M., Yildiz, I. B., Macedonia, M., and von Kriegstein, K. (2015). Visual and motor cortices differentially support the translation of foreign language words. *Current biology*, 25(4), pp. 530–535.

Mazzoni, E. and Benvenuti, M. (2015). A robot-partner for preschool children learning english using socio-cognitive conflict. *Journal of Educational Technology & Society*, 18(4), pp. 474–485.

McColl, D. and Nejat, G. (2012). Affect detection from body language during social hri. In *Proceedings of the 21st IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '12)*, Paris, France, pp. 1013–1018.

McDaniel, B., D'Mello, S., King, B., Chipman, P., Tapp, K., and Graesser, A. (2007). Facial features for affective state detection in learning environments. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society (Cogsci '07)*, volume 29, Nashville, Tennessee, USA.

McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., and Kaliouby, R. e. (2016). Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit. In *Proceedings of the Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '16)*, San Jose, California, USA, pp. 3723–3726. DOI:10.1145/2851581.2890247.

McQuiggan, S. W., Mott, B. W., and Lester, J. C. (2008). Modeling self-efficacy in intelligent tutoring systems: An inductive approach. *User Modeling and User-Adapted Interaction*, 18(1), pp. 81–123. DOI:10.1007/s11257-007-9040-y.

Mechelli, A., Crinion, J., Noppeney, U., O'Doherty, J., Ashburner, J., Frackowiak, R., and Price, C. (2004). Structural pasticity in the bilingual brain: proficiency in a second language and age at acquisition affect grey-matter density. *Nature*, 431, pp. 757–757.

Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barnes, M. J., Barber, D., and Procci, K. (2016). Intelligent agent transparency in human–agent teaming for multi-uxv management. *Human Factors*, 58(3), pp. 401–415. DOI:10.1177/0018720815621206.

Miehle, J., Bagci, I., Minker, W., and Ultes, S. (2019). A social companion and conversational partner for the elderly. In Eskenazi, M., Devillers, L., and Mariani, J., editors, *Advanced Social Interaction with Agents*, pp. 103–109. Springer International Publishing, Cham, Germany.

Miller, R. B., Greene, B. A., Montalvo, G. P., Ravindran, B., and Nichols, J. D. (1996). Engagement in academic work: The role of learning goals, future consequences, pleasing others, and perceived ability. *Contemporary Educational Psychology*, 21(4), pp. 388–422. DOI:10.1006/ceps.1996.0028.

Mitrovic, A. (2003). An intelligent sql tutor on the web. *International Journal of Artificial Intelligence in Education*, 13(2-4), pp. 173–197.

Mitrovic, A. (2010). Modeling domains and students with constraint-based modeling. In Nkambou, R., Bourdeau, J., and Mizoguchi, R., editors, *Advances in intelligent tutoring systems*, pp. 63–80. Springer Berlin Heidelberg, Berlin, Germany.

Mitrovic, A., Martin, B., and Suraweera, P. (2007). Intelligent tutors for all: The constraint-based approach. *IEEE Intelligent Systems*, 22(4), pp. 38–45. DOI:10.1109/MIS.2007.74.

Montemerlo, M., Pineau, J., Roy, N., Thrun, S., and Verma, V. (2002). Experiences with a mobile robotic guide for the elderly. In *Proceedings of the 18th National Conference on Artificial Intelligence (AAAI '02)*, Edmonton, Alberta, Canada, pp. 587–592.

Mooney, C. G. (2013). *Theories of Childhood: An Introduction to Dewey, Montessori, Erikson, Piaget & Vygotsky*. Redleaf Press, St Paul, Minnesota, USA, 2 edition.

Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional science*, 32(1-2), pp. 99–113. DOI:10.1023/B:TRUC.0000021811.66966.1d.

Moritz, S. and Blank, G. (2008). Generating and evaluating object-oriented designs for instructors and novice students. In *Proceedings of the Intelligent Tutoring Systems for Ill-Defined Domains Workshop*, Montreal, Quebec, Canada, pp. 35–45.

Movellan, J. R., Eckhardt, M., Virnes, M., and Rodriguez, A. (2009). Sociable robot improves toddler vocabulary skills. In *Proceedings of the 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI '09)*, La Jolla, California, USA, pp. 307–308. DOI:10.1145/1514095.1514189.

Müller, H. M. (2013). *Psycholinguistik - Neurolinguistik. Die Verarbeitung von Sprache im Gehirn.* UTB GmbH, Stuttgart, Germany.

Naiman, N., Fröhlich, M., Stern, H. H., and Todesco, A. (1996). *The good language learner*. Modern languages in practice: 4. Multilingual Matters, Bristol, Pennsylvania, USA.

Nkambou, R., Bourdeau, J., and Mizoguchi, R. (2010). *Advances in Intelligent Tutoring Systems*. Springer, Heidelberg, Germany.

Nwana, H. S. (1990). Intelligent tutoring systems: An overview. *Artificial Intelligence Review*, 4(4), pp. 251–277. DOI:10.1007/BF00168958.

Olney, A. M., D'Mello, S., Person, N., Cade, W., Hays, P., Williams, C., Lehman, B., and Graesser, A. (2012). Guru: A computer tutor that models expert human tutors. In *Proceedings of the 10th International Conference on Intelligent Tutoring Systems (ITS '12)*, Chania, Crete, Greece, pp. 256–261.

Orejana, J. R., MacDonald, B. A., Ahn, H. S., Peri, K., and Broadbent, E. (2015). Healthcare robots in homes of rural older adults. In *Proceedings of the 7th International Conference on Social Robotics (ICSR '15)*, Paris, France, pp. 512–521.

Oxford, R. L. (2003). *Language learning styles and strategies*. Mouton de Gruyter, Berlin, Germany.

Pajares, F. and Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Educational Psychology*, 86(2), pp. 193–203. DOI:10.1037/0022-0663.86.2.193.

Paleari, M., Lisetti, C., and Lethonen, M. (2005). VALERIE: a virtual agent for learning environment, reacting and interacting emotionally. In *Proceedings of the 12th International Conference on Artificial Intelligence in Education (AIED '05)*, Amsterdam, The Netherlands, pp. 2–5.

Pardos, Z. A., Gowda, S. M., Baker, R. S., and Heffernan, N. T. (2012). The sum is greater than the parts: Ensembling models of student knowledge in educational software. *SIGKDD Explorations Newsletter*, 13(2), pp. 37–44. DOI:10.1145/2207243.2207249.

Pardos, Z. A. and Heffernan, N. T. (2010). Modeling individualization in a bayesian networks implementation of knowledge tracing. In De Bra, P., Kobsa, A., and Chin, D., editors, *Proceedings of the 11th International Conference on User Modeling, Adaptation, and Personalization (UMAP '10)*, Manoa, Hawaii, USA, pp. 255–266.

Pardos, Z. A. and Heffernan, N. T. (2011). Kt-idem: Introducing item difficulty to the knowledge tracing model. In *Proceedings of the 12th International Conference on User Modeling, Adaptation, and Personalization (UMAP '11)*, Girona, Spain, pp. 243–254.

Pat-El, R. J., Tillema, H., Segers, M., and Vedder, P. (2013). Validation of assessment for learning questionnaires for teachers and students. *British Journal of Educational Psychology*, 83(1), pp. 98–113. DOI:10.1111/j.2044-8279.2011.02057.x.

Paviotti, G., Rossi, P. G., and Zarka, D. (2012). *Intelligent Tutoring Systems: An overview*. Pensa Multimedia, Lecce, Italy.

Pavlik, P. I. and Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: An activation-based model of the spacing effect. *Cognitive Science*, 29(4), pp. 559–586. DOI:10.1207/s15516709cog0000_14.

Pavlik, P. I., Cen, H., and Koedinger, K. R. (2009). Performance factors analysis – a new alternative to knowledge tracing. In *Proceedings of the 14th International Conference on Artificial Intelligence in Education (AIED '09)*, Brighton, UK, pp. 531–538.

Peace Corps (1986). *Teacher Training: A Reference Manual.* Peace Corps – Information Collection and Exchange Office of Training and Program Support, Washington, D.C., USA.

Pelánek, R. (2014). Application of time decay functions and the elo system in student modeling. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM '14)*, London, UK, pp. 21–27.

Pelánek, R. (2015). Modeling students' memory for application in adaptive educational systems. In *Proceedings of the 8th International Conference on Educational Data Mining (EDM '15)*, Madrid, Spain, pp. 480–483.

Pelánek, R. (2017). Bayesian knowledge tracing, logistic models, and beyond: an overview of learner modeling techniques. *User Modeling and User-Adapted Interaction*, 27(3), pp. 313–350. DOI:10.1007/s11257-017-9193-2.

Pelánek, R. (2016). Applications of the elo rating system in adaptive educational systems. *Computers and Education*, 98, pp. 169–179. DOI:10.1016/j.compedu.2016.03.017.

Peters, C., Pelachaud, C., Bevacqua, E., Mancini, M., Poggi, I., and Tre, U. R. (2005). Engagement capabilities for ECAs. In *Proceedings of the Creating Bonds with ECAs Workshop at the 4th international conference on autonomous agents & multi agent systems (AAMAS '05)*, Utrecht, The Netherlands.

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., and Sohl-Dickstein, J. (2015). Deep knowledge tracing. In Cortes, C., Lawrence, N. D., Lee, D. D., Sugiyama, M., and Garnett, R., editors, *Advances in Neural Information Processing Systems 28*, pp. 505–513. Curran Associates Inc., Red Hook, New York, USA.

Pineau, J., Montemerlo, M., Pollack, M., Roy, N., and Thrun, S. (2003). Towards robotic assistants in nursing homes: Challenges and results. *Robotics and autonomous systems*, 42(3-4), pp. 271–281.

Polich, K. and Gmytrasiewicz, P. (2007). Interactive dynamic influence diagrams. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS '07)*, Honolulu, Hawaii, pp. 34:1–34:3. DOI:10.1145/1329125.1329166.

Psotka, J., Massey, L. D., and Mutter, S. A., editors (1988). *Intelligent Tutoring Systems: Lessons Learned.* Lawrence Erlbaum Associates, Inc, Hillsdale, New Jersey, USA.

Qiu, Y., Qi, Y., Lu, H., Pardos, Z. A., and Heffernan, N. T. (2011). Does time matter? modeling the effect of time with bayesian knowledge tracing. In *Proceedings of the 4th International Conference on Educational Data Mining (EDM '11)*, Einhoven, The Netherlands, pp. 139–148.

Rafferty, A. N., Brunskill, E., Griffiths, T. L., and Shafto, P. (2011). Faster teaching by pomdp planning. In *Proceedings of 15th International Conference on Artificial Intelligence in Education (AIED '11)*, Auckland, New Zealand, pp. 280–287.

Ramachandran, A., Huang, C.-M., Gartland, E., and Scassellati, B. (2018). Thinking aloud with a tutoring robot to enhance learning. In *Proceedings of the 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*, Chicago, Illinois, USA, pp. 59–68. DOI:10.1145/3171221.3171250.

Ramachandran, A., Huang, C.-M., and Scassellati, B. (2017). Give me a break!: Personalized timing strategies to promote learning in robot-child tutoring. In *Proceedings of the 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*, Vienna, Austria, pp. 146–155. DOI:10.1145/2909824.3020209.

Ramachandran, A. and Scassellati, B. (2015). Fostering learning gains through personalized robot-child tutoring interactions. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts (HRI EA '15)*, Portland, Oregon, USA, pp. 193–194. DOI:10.1145/2701973.2702721.

Ramirez, N. F. and Kuhl, P. K. (2016). Bilingual language learning in children. *Institute of Learning & Brain Science*.

Rintjema, E., van den Berghe, R., Kessels, A., de Wit, J., and Vogt, P. (2018). A robot teaching young children a second language: The effect of multiple interactions on engagement and performance. In *Proceedings of the Companion of the 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI '18)*, Chicago, Illinois, USA, pp. 219–220. DOI:10.1145/3173386.3177059.

Ritter, S., Anderson, J. R., Koedinger, K. R., and Corbett, A. (2007). Cognitive tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, 14(2), pp. 249–255. DOI:10.3758/BF03194060.

Robins, B., Dautenhahn, K., and Dickerson, P. (2012). Embodiment and cognitive learning – can a humanoid robot help children with autism to learn about tactile social behaviour? In *Proceedings of the 3rd International Conference of Social Robotics (ICSR '12)*, Chengdu, China, pp. 66–75.

Robins, B., Dautenhahn, K., Te Boekhorst, R., and Billard, A. (2005). Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? *Universal Access in the Information Society*, 4(2), pp. 105–120.

Rohlfing, K. J., Wrede, B., Vollmer, A.-L., and Oudeyer, P.-Y. (2016). An alternative to mapping a word onto a concept in language acquisition: pragmatic frames. *Frontiers in psychology*, 7, pp. 470–487. DOI:10.3389/fpsyg.2016.00470.

Roseberry, S., Hirsh-Pasek, K., Parish-Morris, J., and Golinkoff, R. M. (2009). Live action: Can young children learn verbs from video? *Child Development*, 80(5), pp. 1360–1375.

Rosenthal-von der Pütten, A. M., Straßmann, C., and Krämer, N. C. (2016). Robots or agents – neither helps you more or less during second language acquisition. In *Proceedings of the 14th International Conference on Intelligent Virtual Agents*, Los Angeles, California, USA, pp. 256–268.

Rowe, M. L. and Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, 323(5916), pp. 951–953.

Rudovic, O., Lee, J., Dai, M., Schuller, B., and Picard, R. W. (2018). Personalized machine learning for robot perception of affect and engagement in autism therapy. *Science Robotics*, 3(19). DOI:10.1126/scirobotics.aao6760.

Russell, S. J. and Norvig, P. (2010). *Artificial Intelligence - A Modern Approach*. Pearson Education, Inc., Upper Saddle River, New Jersey, USA.

Šabanović, S., Bennett, C. C., Chang, W.-L., and Huber, L. (2013). Paro robot affects diverse interaction modalities in group sensory therapy for older adults with dementia. In *Proceedings of the 13th IEEE International Conference on Rehabilitation Robotics (ICORR '13)*, Seattle, Washington, USA, pp. 1–6. DOI:10.1109/ICORR.2013.6650427.

Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2), pp. 159–179.

Saerbeck, M., Schut, T., Bartneck, C., and Janse, M. D. (2010). Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '10)*, Atlanta, Georgia, USA, pp. 1613–1622. DOI:10.1145/1753326.1753567.

Saldien, J., Goris, K., Vanderborght, B., Verrelst, B., Van Ham, R., and Lefeber, D. (2006). Anty: The development of an intelligent huggable robot for hospitalized children. In *Proceedings of the 9th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines (CLAWAR '06)*, Coimbra, Portugal, pp. 123–128.

Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P. W., and Paiva, A. (2011). Automatic analysis of affective postures and body motion to detect engagement with a game companion. In *Proceedings of the 6th International Conference on Human-robot Interaction (HRI '11)*, Lausanne, Switzerland, pp. 305–312. DOI:10.1145/1957656.1957781.

Sarma, B. H. S. and Ravindran, B. (2007). Intelligent tutoring systems using reinforcement learning to teach autistic students. In *Proceedings of the 2nd International Conference on Home-Oriented Informatics and Telematics (HOIT '07)*, Chennai, India, pp. 65–78.

Schadenberg, B., Neerincx, M., Cnossen, F., and Looije, R. (2017). Personalising game difficulty to keep children motivated to play with a social robot: A Bayesian approach. *Cognitive Systems Research*, 43, pp. 222–231. DOI:10.1016/j.cogsys.2016.08.003.

Schlegel, A. A., Rudelson, J. J., and Tse, P. U. (2012). White matter structure changes as adults learn a second language. *Journal of Cognitive Neuroscience*, 24(8), pp. 1664–1670.

Schodde, T., Bergmann, K., and Kopp, S. (2017a). Adaptive robot language tutoring based on bayesian knowledge tracing and predictive decision-making. In *Proceedings of the 12th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '17)*, Vienna, Austria, pp. 128–136. DOI:10.1145/2909824.3020222.

Schodde, T., Hoffmann, L., and Kopp, S. (2017b). How to manage affective state in child-robot tutoring interactions? In *Proceedings of the 2nd International Conference on Companion Technology (ICCT '17)*, Ulm, Germany, pp. 1–6. DOI:10.1109/COMPANION.2017.8287073.

Schodde, T., Hoffmann, L., Stange, S., and Kopp, S. (2019). Adapt, explain, engage – A study on how social robots can scaffold second-language learning of children. Manuscript submitted for publication.

Schunk, D. H. (1987). Peer models and children's behavioral change. *Review of Educational Research*, 57(2), pp. 149–174. DOI:10.3102/00346543057002149.

Schunk, D. H. and Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading & Writing Quarterly*, 23(1), pp. 7–25. DOI:10.1080/10573560600837578.

Schwarz, N. (2000). Emotion, cognition, and decision making. *Cognition and Emotion*, 14(4), pp. 433–440. DOI:10.1080/026999300402745.

Senft, E., Lemaignan, S., Bartlett, M., Baxter, P., Belpaeme, T., et al. (2018). Robots in the classroom: Learning to be a good tutor. In *Proceedings of the Robots for Learning Workshop at the 13th International Conference of Human-Robot Interaction (R4L '18)*, Chicago, Illinois, USA.

Sense, F., Behrens, F., Meijer, R. R., and Rijn, H. (2016). An individual's rate of forgetting is stable over time but differs across materials. *Topics in Cognitive Science*, 8(1), pp. 305–321. DOI:10.1111/tops.12183.

Serholt, S., Barendregt, W., Ribeiro, T., Castellano, G., Paiva, A., Kappas, A., Aylett, R., and Nabais, F. (2013). Emote: Emboided-perceptive tutors for empathy-based learning in game environment. In *Proceedings of the 7th European Conference on Games Based Learning (ECGBL '13)*, Porto, UK, pp. 790–792.

Serholt, S., Basedow, C. A., Barendregt, W., and Obaid, M. (2014). Comparing a humanoid tutor to a human tutor delivering an instructional task to children. In *Proceedings of the IEEE-RAS International Conference on Humanoid Robots*, Madrid, Spain, pp. 1134–1141. DOI:10.1109/HUMANOIDS.2014.7041511.

Shahid, S., Krahmer, E., and Swerts, M. (2011). Child-robot interaction: Playing alone or together? In *Extended Abstracts on Human Factors in Computing Systems (CHI EA '11)*, Vancouver, British Columbia, Canada, pp. 1399–1404. DOI:10.1145/1979742.1979781.

Shen, L., Wang, M., and Shen, R. (2009). Affective e-Learning: Using "emotional" data to improve learning in pervasive learning environment related work and the pervasive e-learning platform. *Educational Technology & Society*, 12, pp. 176–189. DOI:citeulike-article-id:7412147.

Shepard, L. A. (2005). Linking formative assessment to scaffolding. *Educational Leadership*, 63(3), pp. 66–70.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), pp. 153–189. DOI:10.3102/0034654307313795.

Siemer, J. and Angelides, M. C. (1998). A comprehensive method for the evaluation of complete intelligent tutoring systems. *Decision Support Systems*, 22(1), pp. 85–102. DOI:10.1016/S0167-9236(97)00033-X.

Simpson, E. J. (1972). *The Classification of Educational Objectives in the Psychomotor Domain*. Gryphon House, Washington, DC, USA.

Simpson, R. C. and Levine, S. P. (1997). Development and evaluation of voice control for a smart wheelchair. In *Proceedings of the Annual Conference of the Rehabilitation Engineering Society of North America (RESNA '97)*, Pittsburgh, Pennsylvania, USA, pp. 417–419.

Singer, I. and Gerrits, E. (2015). The effect of playing with tablet games compared with real objects on word learning by toddlers. In *Proceedings of the 8th International Conference on ICT for Language Learning*, Florence, Italy, pp. 255–259.

Singleton, N. C. (2012). Can semantic enrichment lead to naming in a word extension task? *American Journal of Speech-Language Pathology*, 21(4), pp. 279–292.

Skinner, E. A. and Belmont, M. J. (1993). Motivation in the classroom: Reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, 85(4), pp. 571–581.

So, W.-C., Wong, M. K.-Y., Lam, W.-Y., Cheng, C.-H., Ku, S.-Y., Lam, K.-Y., Huang, Y., and Wong, W.-L. (2019). Who is a better teacher for children with autism? comparison of learning outcomes between robot-based and human-based interventions in gestural production and recognition. *Research in Developmental Disabilities*, 86, pp. 62–75. DOI:10.1016/j.ridd.2019.01.002.

Sparks, R. L. (2012). Individual differences in L2 learning and long-term L1–L2 relationships. *Language Learning*, 62(s2), pp. 5–27. DOI:10.1111/j.1467-9922.2012.00704.x.

Spaulding, S., Gordon, G., and Breazeal, C. (2016). Affect-aware student models for robot tutors. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '16)*, Singapore, pp. 864–872.

Spinath, B., Spinath, F. M., Harlaar, N., and Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence*, 34(4), pp. 363–374. DOI:10.1016/j.intell.2005.11.004.

Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., and Samson, A. C. (2018). Facial expression analysis with affdex and facet: A validation study. *Behavior Research Methods*, 50(4), pp. 1446–1460.

Streeter, M. (2015). Mixture modeling of individual learning curves. In *Proceedings of the 8th International Conference on Educational Data Mining (EDM '15)*, Madrid, Spain, pp. 45–52.

Stroud, P., Jones, R., and Brien, S. (2018). Global people movements - a report published by the legatum institute foundation in partnership with oxford analytica. https://lif.blob.core.windows.net/lif/docs/default-source/default-library/legj6267_global-people-movements-180622.pdf?sfvrsn=0.

Suraweera, P., Mitrovic, A., and Martin, B. (2005). A knowledge acquisition system for constraint-based intelligent tutoring systems. In *Proceedings of the 12th International Conference on Artificial Intelligence in Education (AIED '12)*, Amsterdam, The Netherlands, pp. 638–645.

Szafir, D. and Mutlu, B. (2012). Pay attention!: Designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*, Austin, Texas, USA, pp. 11–20. DOI:10.1145/2207676.2207679.

Sénéchal, M. (1997). The differential effect of storybook reading on preschoolers' acquisition of expressive and receptive vocabulary. *Journal of Child Language*, 24(1), pp. 123–138.

Tanaka, F. and Matsuzoe, S. (2012a). Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction*, 1(1), pp. 78–95. DOI:10.5898/JHRI.1.1.Tanaka.

Tanaka, F. and Matsuzoe, S. (2012b). Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction*, 1(1), pp. 78–95.

Tapus, A., Mataric, M. J., and Scassellati, B. (2007). Socially assistive robotics: Grand challenges of robotics. *IEEE Robotics Automation Magazine*, 14(1), pp. 35–42. DOI:10.1109/MRA.2007.339605.

Theocharous, G. (2010). Designing a mathematical manipulatives tutoring system using pomdps. In *Proceedings at the POMDP practitioners Workshop on Solving Real-world POMDP Problems at the 20th international Conference on Automated Planning and Scheduling (ICAPS '10)*, Toronto, Ontario, Canada.

Töscher, A. and Jahrer, M. (2010). Collaborative filtering applied to educational data mining. In *Proceedings of the KDD Cup 2010 Workshop: Knowledge discoveryin educational data*, Washington, DC, USA, pp. 13–23.

Toumpaniari, K., Loyens, S., Mavilidi, M.-F., and Paas, F. (2015). Preschool children's foreign language vocabulary learning by embodying words through physical activity and gesturing. *Educational Psychology Review*, 27(3), pp. 445–456. DOI:10.1007/s10648-015-9316-4.

Tyng, C. M., Amin, H. U., Saad, M. N., and Malik, A. S. (2017). The influences of emotion on learning and memory. *Frontiers in psychology*, 8, pp. 1454. DOI:10.3389/fpsyg.2017.01454.

Tzirakis, P., Zhang, J., and Schuller, B. W. (2018). End-to-end speech emotion recognition using deep neural networks. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '18)*, Calgary, Alberta, Canada, pp. 5089–5093.

Uluer, P., Akalın, N., and Köse, H. (2015). A new robotic platform for sign language tutoring. *International Journal of Social Robotics*, 7(5), pp. 571–585.

van den Berghe, R., Verhagen, J., Oudgenoeg-Paz, O., van der Ven, S., and Leseman, P. (2019). Social robots for language learning: A review. *Review of Educational Research*, 89(2), pp. 259–295. DOI:10.3102/0034654318821286.

Vanderborght, B., Simut, R., Saldien, J., Pop, C., Rusu, A. S., Pintea, S., Lefeber, D., and David, D. O. (2012). Using the social robot probo as a social story telling agent for children with asd. *Interaction Studies*, 13(3), pp. 348–372.

VanLehn, K. (2011). The relative effectiveness of human tutoring , intelligent tutoring systems , and other tutoring systems. *Educational Psychologist*, 46(4), pp. 197–221.

Vanlehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., Treacy, D., Weinstein, A., and Wintersgill, M. (2005). The andes physics tutoring system: Lessons learned. *Artificial Intelligence in Education*, 15(3), pp. 147–204.

Vanlehn, K. and Martin, J. (1998). Evaluation of an assessment system based on bayesian student modeling. *International Journal of Artificial Intelligence in Education*, 8, pp. 179–221.

Villon, O. and Lisetti, C. (2006). A user-modeling approach to build user's psycho-physiological maps of emotions using bio-sensors. In *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '06)*, Hatfield, UK, pp. 269–276. DOI:10.1109/ROMAN.2006.314429.

Vlaar, R., Verhagen, J., Oudgenoeg-Paz, O., and Leseman, P. (2017). Comparing L2 word learning through a tablet or real objects: What benefits learning most? In *Proceedings of the Robots for Learning Workshop at the 12th Annual International Conference of Human-Robot Interaction (R4L '17)*, Vienna, Austria.

Vogt, P., Haas, M. D., Jong, C. D., Baxter, P., and Krahmer, E. (2017). Child-robot interactions for second language tutoring to preschool children. *Frontiers in Human Neuroscience*, 11(March), pp. 1–7. DOI:10.3389/fnhum.2017.00073.

Vogt, P., van den Berghe, R., de Haas, M., Hoffmann, L., Kanero, J., Mamus, E., Montanier, J.-M., Oudgenoeg-Paz, O., Garcia, D. H., Papadopoulos, F., et al. (2019). Second language tutoring using social robots. a large-scale study. In *Proceesdings of the 14th Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '19)*, Daegu, Korea, pp. 497–505. DOI:10.1109/HRI.2019.8673077.

Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press, Cambridge, UK.

Wagner, J., Kim, J., and André, E. (2005). From physiological signals to emotions: Implementing and comparing selected methods for feature extraction and classification. In *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '05)*, Amsterdam, The Netherlands, pp. 940–943. DOI:10.1109/ICME.2005.1521579.

Wainer, J., Feil-Seifer, D. J., Shell, D. A., and Mataric, M. J. (2007). Embodiment and human-robot interaction: A task-based perspective. In *Proceedings of the 16th IEEE International Symposium on Robot and Human Interactive Communication, ROMAN 2007*, Jeju Island, Korea, pp. 872–877. DOI:10.1109/ROMAN.2007.4415207.

Wainer, J., Robins, B., Amirabdollahian, F., and Dautenhahn, K. (2014). Using the humanoid robot kaspar to autonomously play triadic games and facilitate collaborative play among children with autism. *IEEE Transactions on Autonomous Mental Development*, 6(3), pp. 183–199. DOI:10.1109/TAMD.2014.2303116.

Wang, S.-H., Phillips, P., Dong, Z.-C., and Zhang, Y.-D. (2018). Intelligent facial emotion recognition based on stationary wavelet entropy and jaya algorithm. *Neurocomputing*, 272, pp. 668 – 676. DOI:10.1016/j.neucom.2017.08.015.

Wang, X., Berger, J. O., and Burdick, D. S. (2013). Bayesian analysis of dynamic item response models in educational testing. *The Annals of Applied Statistics*, 7(1), pp. 126–153. DOI:10.1214/12-AOAS608.

Wauters, K., Desmet, P., and Noortgate, W. V. D. (2012). Item difficulty estimation: An auspicious collaboration between data and judgment. *Computers and Education*, 58(4), pp. 1183–1193. DOI:10.1016/j.compedu.2011.11.020.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology*, 71(1), pp. 3–25.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological review*, 92(4), pp. 548–573.

Welch, L. R. (2003). Hidden markov models and the baum-welch algorithm. *IEEE Information Theory Society Newsletter*, 53(4), pp. 9–13.

Westlund, J. K., Dickens, L., Jeong, S., Harris, P., DeSteno, D., and Breazeal, C. (2015). A comparison of children learning new words from robots, tablets, & people. In *Proceedings of the 1th International Conference on New Friends*, Almere, The Netherlands, pp. 26–27.

Westlund, J. M. K., Dickens, L., Jeong, S., Harris, P. L., DeSteno, D., and Breazeal, C. L. (2017). Children use non-verbal cues to learn new words from robots as well as people. *International Journal of Child-Computer Interaction*, 13, pp. 1–9.

White, K. G. (2001). Forgetting functions. *Animal Learning & Behavior*, 29(3), pp. 193–207. DOI:10.3758/BF03192887.

Wixon, M., Arroyo, I., Muldner, K., Burleson, W., Rai, D., and Woolf, B. (2014). The opportunities and limitations of scaling up sensor-free affect detection. In *Proceedings of the 7th International Conference on Educational Data Mining (EDM '14)*, London, UK, pp. 145–152.

Wood, D., Bruner, J. S., and Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, 17(2), pp. 89–100.

Yang, F.-Y., Chang, C.-Y., Chien, W.-R., Chien, Y.-T., and Tseng, Y.-H. (2013). Tracking learners' visual attention during a multimedia presentation in a real classroom. *Computers & Education*, 62, pp. 208–220. DOI:10.1016/j.compedu.2012.10.009.

Yang, X., Shyu, M., Yu, H., Sun, S., Yin, N., and Chen, W. (2019). Integrating image and textual information in human–robot interactions for children with autism spectrum disorder. *IEEE Transactions on Multimedia*, 21(3), pp. 746–759. DOI:10.1109/TMM.2018.2865828.

You, Z.-J., Shen, C.-Y., Chang, C.-W., Liu, B.-J., and Chen, G.-D. (2006). A robot as a teaching assistant in an english class. In *Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies (ICALT '06)*, Kerkrade, The Netherlands, pp. 87–91.

Yudelson, M. V., Koedinger, K. R., and Gordon, G. J. (2013). Individualized bayesian knowledge tracing models. In *Proceedings of the 2nd International Conference on Artificial Intelligence in Education (AIED '13)*, Memphis, Tennessee, USA, pp. 171–180.

Zaraki, A., Wood, L., Robins, B., and Dautenhahn, K. (2018). Development of a semi-autonomous robotic system to assist children with autism in developing visual perspective taking skills. In *Proceedings of the 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '18)*, Nanjing, China, pp. 969–976.

Zhang, K. and Yao, Y. (2018). A three learning states bayesian knowledge tracing model. *Knowledge-Based Systems*, 148, pp. 189 – 201. DOI:10.1016/j.knosys.2018.03.001.

Zhu, J., Zang, Y., Qiu, H., and Zhou, T. (2018). Integrating temporal information into knowledge tracing: A temporal difference approach. *IEEE Access*, 6, pp. 27302–27312. DOI:10.1109/ACCESS.2018.2833874.

Zimmerman, B. J. and Kitsantas, A. (1996). Self-regulated learning of a motoric skill: The role of goal setting and self-monitoring. *Journal of Applied Sport Psychology*, 8(1), pp. 60–75. DOI:10.1080/10413209608406308.

Zimmerman, B. J. and Kitsantas, A. (1997). Developmental phases in self-regulation: Shifting from process goals to outcome goals. *Journal of Educational Psychology*, 89(1), pp. 29–36.

Affidavit

Hiermit erkläre ich, dass ich die vorliegende Dissertation konform zu § 8 Abs. 1 lit g der Rahmenpromotionsordnung der Universität Bielefeld vom 15. Juni 2010[1] angefertigt habe. Dies bedeutet, dass

- mir die geltende Promotionsordnung der Technischen Fakultät der Universität Bielefeld vom 1. März 2011[2] bekannt ist;
- ich die Dissertation selbst angefertigt, keine Textabschnitte von Dritten oder eigenen Prüfungsarbeiten ohne Kennzeichnung übernommen und alle von mir benutzten Hilfsmittel und Quellen in meiner Arbeit angegeben habe;
- Dritte weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Vermittlungstätigkeiten oder für Arbeiten erhalten haben, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen;
- diese Dissertation noch nicht als Prüfungsarbeit für eine staatliche oder andere wissenschaftliche Prüfung eingereicht wurde;
- die gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung von mir bei keiner anderen Hochschule als Dissertation eingereicht wurde.

Bielfeld, December 9, 2019

Thorsten Schodde

[1] In: *Verkündungsblatt—Amtliche Bekanntmachungen.* Ed. by Rektorat der Universität Bielefeld. Vol. 39. Bielefeld, Germany: Universität Bielefeld, pp. 98–105

[2] Idem. Vol. 40, pp. 56–59