




Rebo Junior: Analysis of Dialogue Structure Quality for a Reflection Guidance Chatbot

Irmtraud Wolfbauer¹ , Viktoria Pammer-Schindler^{1,2}  and Carolyn P. Rose³ 

¹ Know-Center GmbH, Inffeldgasse 13, 8010 Graz, Austria
iwolfbauer@know-center.at  <https://orcid.org/0000-0002-2973-9680>

² Graz University of Technology, 8010 Graz, Austria
viktoria.pammer@tugraz.at  <https://orcid.org/0000-0001-7061-8947>

³ Carnegie Mellon University, Pittsburgh PA 15213, USA
cprose@cs.cmu.edu  <https://orcid.org/0000-0003-1128-5155>

Abstract. Conversational user interfaces open up new opportunities for reflection guidance. This paper presents a computer-mediated dialogue structure for reflecting on learning tasks, Rebo Junior, and its evaluation in the context of apprenticeship training. We answer three research questions. Firstly, how apprentices react to Rebo Junior; secondly, whether Rebo Junior’s dialogue structure is apt to lead apprentices in reflective conversations; and thirdly, how user engagement with Rebo Junior develops over time. During three months, 17 apprentices led 153 reflective conversations with Rebo Junior in the context of a training workshop, 117 in phase one and 36 in phase two of the study (five to thirteen interactions per apprentice). We coded interactions manually for coherence, level of reflectivity, and user engagement. Our results show that apprentices react well to the intervention and that the dialogue structure is successful in leading apprentices through different levels of reflection (114 out of 153 showed observable reflection on the learning experience; 133 out of 153 expressed learning or planned behaviour change for future tasks). Furthermore, the interactions between the apprentices and Rebo Junior result in coherent conversations (149 out of 153 were coherent). Contrary to expectations, engagement did not decrease over time in either phase. With the present paper, we therefore publish a dialogue structure for reflecting on learning tasks that has worked extremely well despite no adaptivity in the conversational interface. Overall, we interpret the results of our work as underscoring the importance of dialogue structure quality in conversational agents.

Keywords: learning technology; reflection guidance; dialogue structure; levels of reflection; reflection guidance chatbot; proof of concept evaluation

1 Introduction and Learning Context

One-on-one reflection with trainers and teachers is unchallenged and not replaceable with conversational agents. However, time with instructors is limited and expensive. For remote locations or in times of quarantine where schools, workshops and factories are closed, it can furthermore be difficult to arrange.

In this paper we report on the development and evaluation of a computer-mediated dialogue structure for reflecting on learning tasks. The computer-mediated dialogue structure was designed as pre-cursor to an adaptive conversational agent, for whom the reported dialogue structure will act as a default dialogue path. The computer-mediated dialogue structure is called “Rebo Junior”. We understand it as the junior version of the future reflection chatbot “Rebo” because it follows through with its pre-defined questions and does not react to the user’s responses.

We have developed and evaluated Rebo Junior in the context of a training workshop for apprentices in electrical engineering, metal and mechatronics. Apprenticeship training for these vocational fields in Austria (similar to Germany and Switzerland) is structured into four years of dual education. Apprentices learn their craft in companies educating apprentices supervised by dedicated apprenticeship supervisors and receive theoretical education at vocational school for a minimum of five weeks each year. The training workshop we collaborate with is a learning site specially financed by participating organisations in addition to obligatory vocational school. In this training workshop, the goal is to teach apprentices fundamental practical knowledge and skills they will need in their workplaces, as well as to provide them with fundamental theoretical knowledge, forging links between theory and application. In each year of apprenticeship training, a pre-defined time is spent in the workshop. The field study described in this paper was conducted with first-year apprentices, who receive a three-month training at the training workshop before starting to work at their respective companies.

Within this training workshop, apprentices receive learning tasks from their trainers. These learning tasks are designed to correspond to the apprentices’ currently expected level of skill and to resemble future workplace tasks. In this learning context, it is the role of Rebo Junior to reflect with each apprentice individually after each practical learning task on how the task went as well as on insights gained and lessons learned for the future. The goals of reflection are to support learning in the domain and to help students improve their ability to reflect, which is considered an important competence in lifelong learning. An example of a learning task is the following: “Produce the workpiece according to the plan. Pay attention to measurements and timing. All measurements in the plan are assessed according to given general tolerance and deviation thereof. (Plans of how to cut materials and assemble them into a pyramid attached)”.

In the field study described here, we evaluate the concept of Rebo, the reflection guidance chatbot and study user engagement and dialogue structure quality. This paper contributes to the existing body of literature evidence about user engagement with a non-adaptive computational dialogue structure throughout 12 weeks (5-13 interactions per apprentice); and a dialogue structure for reflecting on learning tasks.

2 Related Work

2.1 Reflection

By reflection we mean systematic review of past experiences with the goal to learn [1]. Reflection works on different levels (Table 1): Learners remember an experience and think about it carefully. Perceived emotions are attended to [1], the learning experience

is pondered and evaluated, and eventually, the focus is rearranged from retrospective to the future. Learners identify the implications of the experience on future planning and gain new perspectives that, in some cases, affect personal concepts and goals [2–4]. In formal learning environments, reflection helps students to monitor and direct their own learning [1]. In informal learning environments, such as working environments, reflection helps learners to learn from and in relationship to ongoing experience without a dedicated teacher. This emphasizes the importance of reflection for learning for professionals [5]. Constructive, goal-driven reflection is a deliberate action [1] that can be facilitated by reflection guidance technologies [6–8].

Table 1. Levels of Reflection

Level	Name	Description
0	Revisiting	Returning to learning experience
1	Description	Describing learning experience
2	Judgement	Did it go well? Why/Why not?
3	Emotions	What did it feel like? Did you enjoy it?
4	Learning	Evaluate experience - What did you learn?
5	Planning	Behaviour change – And next time?

In the context of our use case, reflection serves as a means for apprentices to engage in a guided manner with past experiences, such as their theoretical and practical lessons as well as their implementation of learning tasks. We want to improve these learning experiences through reflection. Additionally, guided reflection is intended as training in reflection as an important mechanism for lifelong professional learning.

2.2 Conversational Agents for Learning and Reflection Guidance

In comparison to existing literature on reflection guidance, the dialogue structure presented here is new as it provides guidance through different levels of reflection, whereas prior literature has focussed on studying isolated reflection prompts [3, 6, 8]. Furthermore, apprentices are practically not represented in current literature on technologies for learning. Apprenticeship training is situated in the overlap between the informal learning environment of workplace learning and the formal, educational setting of vocational school and trainings. Existing studies on computer-mediated reflection focus on school students [9], university students [7] and professionals [8].

Conversational agents in turn have so far been shown to foster the acquisition of factual knowledge (e.g. [10, 11]), to improve text comprehension by scaffolding self-explanation (e.g. [10, 12]), and to facilitate collaborative learning based on collaboration scripts [13]. However, they have not been used to mediate reflection yet, and they are typically not used in repeated interactions.

In principle, we expect high motivation to reflect with a chatbot because it gives the illusion of a listener [14] and relationships play a critical role in learning [15]. The effect that people prefer interacting with chatbots to other forms of computer-mediated

learning interventions has also been observed in prior research. For example, Ruan et al. [11] showed that students who interact with a dialogue-based agent to acquire factual knowledge displayed more motivation than students interacting with a more traditional computer-mediated learning app. This increased motivation also led to better learning results.

There are, however, very few studies on long-term interactions with conversational agents. Lee et al. [16] created a chatbot to foster self-compassion with which the participants had daily interactions over 2 weeks, which means that each user had 14 interactions with the agent. Upholding user engagement is a key point of interest for chatbot research [17]. On the one hand, it is essential to keep the user interested during the interaction with the agent, as expressed by competitions such as the *Alexa Prize* [18], where keeping users engaged and interested is the goal. On the other hand, the user's engagement has to be upheld over longer timespans when repeated interactions with the agent are planned. With the here presented research, we contribute a field study with repeated chatbot interactions over three months to the existing body of knowledge.

3 Research Questions

We address the following research questions:

RQ1. How do apprentices react to and accept Rebo Junior as reflection guidance?

We understand a positive reaction to a learning intervention as prerequisite for learning [1, 19].

RQ2. How apt is Rebo Junior's dialogue structure to lead reflective conversations with apprentices?

Due to the novelty of conversational reflection guidance, this is a major research question. We understand the suitability of the default dialogue structure to be a prerequisite baseline for an adaptive conversational agent. This needs to be explored in real-world learning contexts within specific and situated reflective conversations.

RQ3. How does apprentices' engagement with Rebo Junior develop over time?

Repeated and long-term interactions with conversational agents are understudied and at the same time, user engagement is crucial for learning. Our initial assumption was that engagement would decrease over time [20].

4 Designing Rebo Junior



Fig. 1. Rebo – the reflection guidance chatbot: This is the design for the image of Rebo, the reflection guidance chatbot. Rebo Junior, the computer-mediated dialogue structure described in this paper, has the same visual appearance as Rebo, since Rebo Junior will successively change into adaptive, more intelligent versions of Rebo.

4.1 Designing Rebo’s Appearance

Rebo’s appearance evolved in a three-cycle iterative design process that aimed to make Rebo engaging and likeable. We aimed for engaging and likeable as these are understood to be prerequisites for users wanting to talk to an agent [21, 22].

Cycle 1. Based on a literature survey, the following initial requirements for Rebo were defined. Rebo needs to look nice and sympathetic, so people want to talk to him. Since social cues were found to be important for motivating users to engage with conversational agents [23] and users tend to prefer visual appearances of chatbots that correspond to the gender stereotypically associated with the task at hand [24], Rebo is referred to as “he”. He needs to look like he is able to communicate (listen, see, talk), but he cannot express emotions because, for instance, a happy face is not suitable for leading a reflective conversation on a bad learning experience. Based on the target audience, Rebo should look cool for young people interested in metal and electronics. Literature suggests that he should not try to appear too human because that could trigger the so-called ‘uncanny valley effect’ in users and make Rebo seem spooky [21]. Based on these requirements, 10 first design ideas for Rebo were sketched out.

Cycle 2. In the second cycle, these ideas were shown to eight people. We settled on the one design that nobody had any objections to, as rejection caused by negative feelings outweighs acceptance by positive reaction [25].



Cycle 3. The starting point was, once again, a literature survey. It was found that people feel more inclined to talk to chatbots if they perceive them as high-quality artefacts [22]. Accordingly, we adapted the design to make it appear high-tech and added some shine and sparkle (Figure 1). The target user group unanimously reacted positively and called Rebo “cool”, so the design was kept.

4.2 Dialogue Design

We have synthesized Boud et al.'s conceptual understanding of reflection as learning mechanism that relates past experiences to future, different experiences [1]; and Fleck & Fitzpatrick's model of different levels of reflection [26] into a hierarchical view of moving through different perspectives on the past towards learning for the future (Table 1). This hierarchical view underlies our design of a dialogue structure.

This dialogue structure is intended to actively guide learners from one level of reflection to the next (Table 2). Our goal is to make sure the learners work through one level after the other. Lower stages of reflection were found to be prerequisites for higher stages in some cases [26], so we want to make sure not to skip a level.

Table 2. Rebo Junior Addresses Subsequent Levels of Reflection

Level	Rebo Junior's Reflective Question	
0 Revisiting	Achieved outside Rebo Junior via upload of task documentation.	
1 Description		
2 Judgement	How was this task for you? Did everything go well?	
3 Emotions	Did you have fun with this task? Why/Why not?	
4 Learning	What tip could you give to a younger apprentice who performs a task like that for the first time?	
5 Planning	What will you pay special attention to when you perform a similar task again?	

The dialogue structure works through the presented levels of reflection as follows. Apprentices return to the experience by accessing the learning platform and viewing the task descriptions before uploading their solutions. The tasks include a description of the performed work or documentation in form of a photograph or video. Therefore, levels 0 and 1 are attended to by uploading the solution to the assigned task and Rebo Junior addresses levels two to five through pre-defined questions (Table 2).

5 Method - Evaluation in a Field Study

5.1 Study Participants

Rebo Junior has been evaluated in a field study with all 18 apprentices in the cohort of 1st year apprentices in the training workshop.

5.2 Procedure – Using Rebo Junior in the Context of a Practical Learning Task

An essential part of apprentices' practical education in the training workshop are learning tasks set by their trainers, where they have to produce a workpiece largely independently and document it digitally (e.g. photograph, video, written documentation).

They upload this documentation to a Moodle-based learning platform¹. Subsequently, the apprentices are directed to Rebo Junior, which is integrated within Moodle, and are guided in reflection on their learning experience.

The apprentices' first interaction with Rebo Junior took place in a workshop setting with one of the authors of this paper. The learning platform was introduced, apprentices worked on their first tasks, documented the completion of their tasks, uploaded task descriptions, and then interacted with Rebo Junior. Directly afterwards, apprentices gave their first reaction and feedback in a focus group as a first measure of reaction (RQ1).

5.3 Repeated Interactions in Two Field Study Phases

The first phase, consisting of the first four weeks of the field study, is characterised by tightly spaced, static interactions with Rebo Junior. All apprentices were present at the training workshop and had daily training where they received practical learning tasks on a regular basis and reflected on them with Rebo Junior. The apprentices had five to nine interactions with Rebo Junior in this phase, 117 altogether.

Phase two, the following eight weeks, is more differentiated and explorative. Interactions with Rebo Junior are more widely spaced because the apprentices received their training in subgroups according to their different professions. Some of these training sequences took place in other locations than the training workshop, where apprentices did not work with the learning platform and with Rebo Junior. In this phase, apprentices also use a version of Rebo Junior that is able to (randomly) vary verbalisations for each reflection level (levels shown in Table 1). Our initial assumption was that engagement would get lower towards the end of phase one and even more so in phase two because repeated interaction with the agent has been found to produce that effect [20]. It has been hypothesised that varying verbalisations are a means to keep up engagement [27]; therefore, we assembled pools of reflective questions for each reflection level and randomly picked a different question for each conversation. These question pools were generated in two workshops, one with colleagues within the research team and one with trainers of the training workshop.

5.4 Data Collection

We collected data by observing the first learning task including apprentices' interactions with Rebo Junior, and in the focus group directly afterwards. Furthermore, we analysed the content of all interactions between the apprentices and Rebo Junior.

5.5 Analysis

All interactions with Rebo Junior were coded for the aspects coherence, reflection depth and engagement. As for coherence as a semantic property of discourses [28], the rationale for using this concept was that interactions with Rebo Junior are intended to be

¹ <https://abvdigital.know-center.tugraz.at>

conversations and therefore have to be coherent. The coding was either 0 (not coherent) or 1 (coherent). As for reflection depth (Table 3), all dialogues were coded according to [29], with 1 (Provision and description of experience), 2 (Reflection on experiences) and 3 (Learning or change). Two researchers coded for coherence and reflectivity, with an inter-rater reliability of 100% for coherence and 97% for reflectivity. We therefore used coherence and reflective depth of recorded conversations as operative measures of how apt Rebo Junior’s dialogue structure is to lead reflective conversations (RQ2). As for engagement (RQ3), we here differentiate between engaged conversations (2), conversations with low engagement (1), where apprentices reacted to Rebo Junior but showed no inclination to be cooperative, and conversations with missing engagement (0), in which Rebo Junior was ignored.

6 Results

The apprentices’ feedback after interacting with Rebo Junior for the first time was captured in a focus group. The results are very positive: 17 out of 18 (94%) apprentices liked interacting with the dialogue structure, and 7 out of 10 who commented on personal gain (70%) see benefit in the guided reflection. Apprentices found that interacting with Rebo Junior “was almost like a real talk²” They commented that it was “really cool that Rebo had a real conversation with you³”. Some apprentices also compared the conversational agent with traditional reflection prompts, such as an empty textbox to fill in or when a teacher gives you a sheet of paper to write down your thoughts, and liked him better. The quality of the following interactions over three months, concerning reflectivity, engagement, as well as the tone of the conversation, further indicate a positive reaction towards and overall acceptance of Rebo Junior.

In the course of our three-month field study, 153 reflective dialogues between the apprentices and Rebo Junior were coded for reflectivity, coherence and engagement. One of the apprentices quit apprenticeship training after their first interaction with Rebo Junior, so we excluded the apprentice from the analysis of the resulting reflective dialogues. The remaining 17 apprentices had 164 interactions with Rebo Junior (between five and 13 per apprentice). 11 interactions had to be removed because of technical problems, so the total number of valid interactions is 153. Of these, 117 are in phase 1, and 36 in phase 2 of the field study.

² Verbatim quote: Ja, fast wie so ein Gespräch mit dir geführt, er hat dich auch so Sachen gefragt. Ja, das war gut!

³ Verbatim quote: Ich habe extrem cool gefunden, dass er so einen richtigen Dialog mit einem geführt hat.

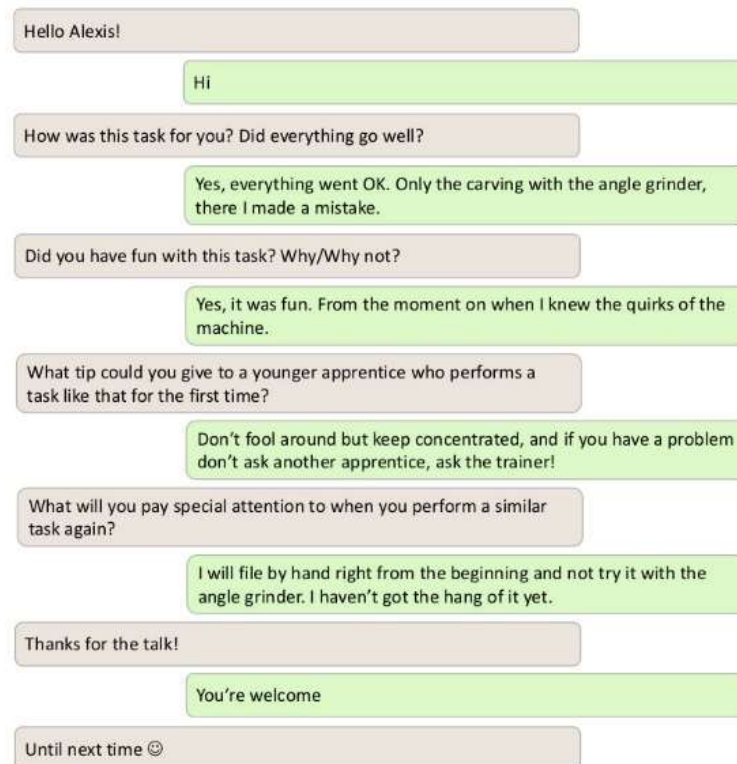


Fig. 2. Coherent and reflective dialogue with Rebo Junior, translated from German. In this dialogue with Rebo Junior, the apprentice successfully reflected on their learning experience.

Figure 2 shows a dialogue which was coded as coherent and highly reflective (levels 2 and 3); the apprentice engages in the conversation, thinks about their learning experience and gives adequate answers. Figure 3 shows a dialogue which was coded as coherent but not reflective (on stages two and three); the apprentice does not really engage in conversation with Rebo Junior but gives very short, non-reflective answers. It could furthermore be observed that with passing time, the answers of apprentices to Rebo Junior's questions are generally getting shorter, sometimes only existing of keywords instead of full sentences, thus less and less resembling human dialogue.

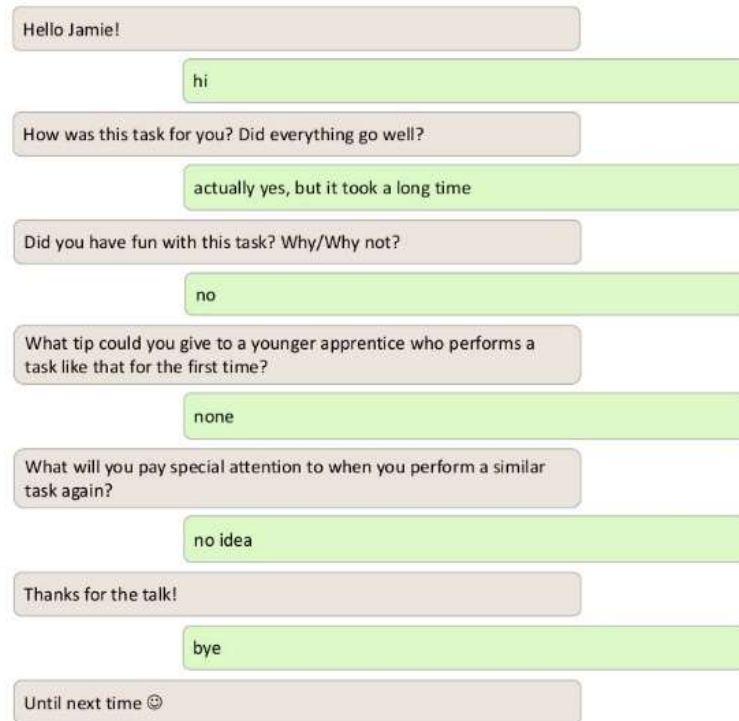


Fig. 3. Dialogue showing missing engagement, translated from German. In this dialogue with Rebo Junior, the apprentice did not successfully reflect on their learning experience because they did not engage in the conversation.

In phase one, nearly all interactions (116 out of 117) were coherent conversations, in the sense of a meaningful sequence of question, answer, and follow-up question. This despite the fact that Rebo Junior, just being a computational interface to a static dialogue structure, does not adapt responses to user statements. The first level of reflection, description, was reached in all interactions because all apprentices needed to upload a description of their learning task prior to reflecting with Rebo Junior. Level two, reflection, was reached in 89 interactions (76%) and level three, learning or change, in 109 interactions (93%) (Table 3). Four interactions (3%) reached only stage one because of missing user engagement. Three out of the four interactions where apprentices did not engage in reflection were still coherent conversations.

Of the 36 valid interactions with Rebo Junior in phase two with randomly picked questions for each level of reflection, 33 were coherent conversations. In the three cases where the resulting dialogue was not coherent, missing engagement was the reason. The first reflection level, description, was reached in all interactions, as explained above. Level two, analysis of the learning experience, was reached in 25 interactions (69%) and level three, learning or change, in 24 interactions (65%) (Table 3). Seven interactions (19%) reached only level one, six of them due to missing user engagement.

Table 3. Coding Interactions between the apprentices and Rebo Junior

Concept	Description	Phase 1 117 int.	Phase 2 36 int.	Overall 153 int.
Coherence*	0: Incoherent, sequence makes no sense	1 (1%)	3 (8%)	4 (3%)
	1: Coherent, given answers and following questions match	116 (99%)	33 (92%)	149 (97%)
Stage of Reflection**	1: Provision and description of experience	117 (100%)	36 (100%)	153 (100%)
	2: Reflection on experiences, including analysis and potential solutions	89 (76%)	25 (69%)	114 (75%)
	3: Learning or change	109 (93%)	24 (65%)	133 (87%)
Engagement	0: Missing engagement	3 (3%)	4 (11%)	7 (5%)
	1: Low engagement	3 (3%)	6 (17%)	9 (6%)
	2: Engaged	111 (95%)	26 (72%)	137 (90%)

* Inter-coder agreement: 100% ** Inter-coder agreement: 97%

For both reflectivity and user engagement, a Chi2 test shows a significant drop in phase two as compared to phase one. The effect is moderate for reflectivity (Chi2=7.680; p=0.021; Cramer's V: 0.232) and considerable for engagement (Chi2= 15.28; p<0.001; Cramer's V= 0.316).

7 Discussion

Rebo Junior is a very successful intervention, in that it has been well received and apprentices to a great extent led reflective conversations with him. In almost all cases, apprentices stayed engaged with Rebo Junior throughout repeated interactions (five to 13 interactions per apprentice over three months), despite the non-adaptiveness of the dialogue structure. This is encouraging for ongoing research on conversational agents for learning, knowing that a positive disposition towards the intervention and continuous engagement is important for learning [1, 17, 19].

Our results also show that the dialogue structure encoded in Rebo Junior successfully facilitates and guides reflection. Those apprentices who engage in a conversation with Rebo Junior are able to reflect on multiple levels, the resulting reflective dialogues throughout portray successful reflection. This validates the quality of the dialogue structure and the initial assumption that engaged learners can be guided by Rebo Junior towards higher levels of reflection.

Despite our initial assumptions, we did not see engagement as decreasing over time when regarding the two phases of the field study separately. Factors positively influencing this continued engagement may be that in repeated interactions the learning task on which apprentices reflect is different every time, designed by their workshop trainers to match the apprentices' current knowledge and skills. It could be that the dialogue

structure working through the levels of reflection in the same order every time was more comforting in its familiarity than boring due to repetition. However, we saw less engagement in interactions with Rebo Junior with different verbalisations.

It is difficult to isolate reasons for the drop in engagement and the lower reflectivity of the resulting dialogues in phase two (varying verbalisations); there are numerous influencing factors. Firstly, the training setting was different than in phase one for a considerable number of apprentices and became altogether more fragmented. Practical learning tasks were scarcer, locations of education varied, and instructors changed for periods outside the training workshop. Secondly, Rebo Junior's question pools were introduced and each conversation varied. Contrary to an earlier voiced assumption [27], such varying verbalisation may have been more impeding by introducing unpredictability than helping by introducing welcome change. Concerning reflectivity, it should also be taken into account that the default dialogue structure was very elaborate and had been developed over weeks, whereas the alternative questions for phase two were generated in a workshop setting with the aim to provide variability. Therefore, it is also possible that not all questions aim as clearly at a specific reflection level while being open enough not to permit single-word answers; in other words, that the concrete dialogue structures in the interactions were simply not as good as in phase one. Overall, we interpret the results concerning the drop in reflectivity as emphasising the importance of careful wording for reflection prompts, especially in conversational reflection guidance.

8 Conclusion

We do not envision conversational reflection guidance to replace human teachers. On the contrary, we fully expect that human teachers will remain unchallenged by technology in principle. However, conditions for human teachers are not always optimal: time is scarce, there are often more students per teacher than would be ideal, and circumstances can prevent teachers and students from getting together. In the kind of vocational settings studied here, the supervisors who train apprentices in their respective companies also have to consider work performance in parallel to apprentices' learning. In all these cases, variants of intelligent tutoring systems may be helpful. Our overall research goal is therefore to develop conversational reflection guidance that can help apprentices to learn how to reflect by pre-structuring reflections, and to learn better within their domain through reflection. As existing computational reflection guidance is mainly based on single prompts or essay writing, conversational guidance is a valuable contribution to the field. Such conversational guidance would in principle be expected to be successful, as natural language conversation is the way humans interact with each other, and especially in reflection, conversations are the traditional way a human teacher would instruct a student. With the present paper, we publish and positively evaluate a dialogue structure for reflecting on learning tasks. Further, we interpret the results of our exploration of varying verbalisation to underscore the importance of exact phrasing to fully exploit dialogue structure quality.

As limitation of the present work, and direction for future work, we see that our analysis of dialogue quality has so far been limited to the dialogic level without considering the content level. In other words, our analysis focusses on what the conversational agent is capable of per design: to structure reflection. This focus of analysis is in line with existing research on reflection analytics [30, 31]. We plan a follow-up study in order to investigate the depth of reflection with respect to correctness and appropriateness of insight within the learning domain, and in relation to the apprentices' expected competence levels. We are especially interested in complementing automatic analyses of reflectivity (reflection analytics) with such a domain-specific dimension. This would have implications for research on reflection analytics, complementing existing research on reflection analytics [30, 31] in two regards: Firstly, extending from analysing reflective essays and statements towards analysing conversations, and secondly, extending from a structural assessment of reflectivity towards including a content-related assessment.

Acknowledgements

This work has partially been funded by the WKO; and within the Austrian COMET Program – Competence Centers for Excellent Technologies – under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth and by the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG. This work was also funded in part by NSF grant IIS 1822831.

References

1. Boud, D., Keogh, R., Walker, D.: Reflection. Turning experience into learning, London, New York (1985).
2. Carrol, M.: Levels of reflection: on learning reflection. *Psychotherapy in Australia* 16 (2010)
3. Renner, B., Prilla, M., Cress, U., Kimmerle, J.: Effects of Prompting in Reflective Learning Tools: Findings from Experimental Field, Lab, and Online Studies. *Frontiers in Psychology* (2016).
4. Wood, D.: Learning from experience through reflection. *Organizational Dynamics* (1996).
5. Pammer, V., Krogstie, B., Prilla, M.: Let's Talk About Reflection at Work. *International Journal of Technology Enhanced Learning* (2015).
6. Ifenthaler, D.: Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios. *Educational Technology & Society* 15, 38–52 (2012).
7. Verpoorten, D., Westera, W. & Specht, M.: Reflection amplifiers in online courses: a classification framework. *Journal of Interactive Learning Research* 2011, 167–190 (22).
8. Fessler, A., Wesiak, G., Rivera-Pelayo, V., Feyertag, S., Pammer, V.: In-App Reflection Guidance: Lessons Learned Across Four Field Trials at the Workplace. *IEEE Trans. Learning Technol.* (2017).
9. Kovanović, V., Joksimović, S., Mirriahi, N., Blaine, E., Gašević, D., Siemens, G., Dawson, S.: Understand students' self-reflections through learning analytics. In: *Proceedings of the 8th Int. Conf. on Learning Analytics & Knowledge*, Sydney. ACM, New York. (2018).

10. Graesser, A. C., VanLehn, K., Rose, C., Jordan, P., Harter, D.: Intelligent Tutoring Systems with Conversational Dialogue. *AI magazine* 22, 39 ff (2001).
11. Ruan, S., Jiang, L., Xu, J., Tham, B.J.-K., Qiu, Z., Zhu, Y., Murnane, E., Brunskill, E., Landay, J.: QuizBot: A Dialogue-based Adaptive Learning System for Factual Knowledge. In: *CHI 2019*, May 4-9, Glasgow, Scotland, UK.
12. Graesser, A.C., McNamara, D.S., VanLehn, K.: Scaffolding Deep Comprehension Strategies Through Point&Query, AutoTutor, and iSTART. *Educational Psychologist* (2005).
13. Adamson, D., Dyke, G., Rosé, C.: Towards an Agile Approach to Adapting Dynamic Collaboration Support to Student Needs. *International Journal of AI in Education* (2014).
14. Knights, S.: Reflection and Learning: The Importance of a Listener. In: Boud et al. (1985). *Reflection: Turning Experience into Learning*, pp. 85–90. RoutledgeFalmer: London, NY.
15. Eraut, M.: Informal learning in the workplace. *Studies in Continuing Education* (2004).
16. Lee, M., Ackermans, S., van As, N., Chang, H., Lucas, E., IJsselsteijn, W.: Caring for Vincent. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*: Glasgow, UK. ACM, New York (2019).
17. Shum, H.-y., He, X.-d., Di Li: From Eliza to XiaoIce: challenges and opportunities with social chatbots. *Frontiers of Information Technology & Electronic Engineering*, 19(1), 10-26 (2018).
18. Alexa Prize, <https://developer.amazon.com/alexaprize> (2020), last accessed 27 April 2020.
19. Kirkpatrick, D.L., Kirkpatrick, J.D.: *Evaluating training programs. The four levels*, 3rd edn. Berrett-Koehler, San Francisco (2010)
20. Kocielnik, R., Xiao, L., Avrahami, D., Hsieh, G.: Reflection Companion. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* (2018).
21. Ciechanowski, L., Przegalinska, A., Magnuski, M., Gloor, P.: In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems* (2019).
22. Zamora, J.: I'm Sorry, Dave, I'm Afraid I Can't Do That. In: *Proceedings of the 5th Int. Conf. on Human Agent Interaction*, Bielefeld, pp. 253–260. ACM Press, New York. (2017).
23. Feine, J., Morana, S., Maedche, A.: Designing a Chatbot Social Cue Configuration System. In: *40th International Conference on Information Systems* (2019).
24. Zimmerman, J., Forlizzi, J., Evenson, S.: Research through design as a method for interaction design research in HCI, pp. 493–502. ACM, New York (2007)
25. Baumeister, R.F., Bratslavsky, E., Finkenauer, C., Vohs, K.D.: Bad is stronger than good. *Review of General Psychology* (2001).
26. Fleck, R., Fitzpatrick, G.: Reflecting on reflection. In: *Proceedings of the 22nd Conf. of the Computer-Human Interaction Special Interest Group of Australia*. ACM, New York (2010).
27. Kocielnik, R., Hsieh, G.: Send Me a Different Message: Utilizing Cognitive Space to Create Engaging Message Triggers. In: *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (2017)
28. van Dijk, T.: *Text and Context. Explorations in the Semantics and Pragmatics of Discourse*. Longman Linguistics Library (1977).
29. Prilla, M., Renner, B.: Supporting Collaborative Reflection at Work. In: *Proceedings of the 18th ACM Int. Conf. on Supporting Group Work*. 182–193. ACM, New York (2014).
30. Cui, Y., Wise, A.F., Allen, K.L.: Developing reflection analytics for health professions education: A multi-dimensional framework to align critical concepts with data features. *Computers in Human Behavior* (2019).
31. Ullmann, T.D.: Automated Analysis of Reflection in Writing: Validating Machine Learning Approaches. *International Journal of Artificial Intelligence in Education* (2019).