In: CHI EA '20: Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu HI USA, April 2020; New York : Association for Computing Machinery; Paper No.: LBW349, https://doi.org/10.1145/3334480.3382805; ISBN 978-1-4503-6819-3



# A Conversational Agent to Improve Response Quality in Course Evaluations

**Thiemo Wambsganss**
University of St.Gallen
St.Gallen, Switzerland
thiemo.wambsganss@unisg.ch
**Rainer Winkler**
University of St.Gallen
St.Gallen, Switzerland
rainer.winkler@unisg.ch
**Matthias Söllner**
University of Kassel
Kassel, Germany
University of St.Gallen
St.Gallen, Switzerland
soellner@uni-kassel.de
**Jan Marco Leimeister**
University of St.Gallen
St.Gallen, Switzerland
University of Kassel
Kassel, Germany
janmarco.leimeister@unisg.ch


*CHI '20 Extended Abstracts, April 25–30, 2020, Honolulu, HI, USA.*

ACM ISBN 978-1-4503-6819-3/20/04.
DOI: https://doi.org/10.1145/3334480.3382805

## Abstract
Recent advances in Natural Language Processing (NLP) bear the opportunity to design new forms of human-computer interaction with conversational interfaces. We hypothesize that these interfaces can interactively engage students to increase response quality of course evaluations in education compared to the common standard of web surveys. Past research indicates that web surveys come with disadvantages, such as poor response quality caused by inattention, survey fatigue or satisficing behavior. To test if conversational interfaces have a positive impact on the level of enjoyment and the response quality, we design an NLP-based conversational agent and deploy it in a field experiment with 127 students in our lecture and compare it with a web survey as a baseline. Our findings indicate that using conversational agents for evaluations are resulting in higher levels of response quality and level of enjoyment, and are therefore, a promising approach to increase the effectiveness of surveys in general.

## Author Keywords
Conversational Agents; Online Course Evaluations; Field Experiment

## Introduction
Web surveys have developed as the standard format for course evaluations in most educational institutions since

**Overview of Hypothesis**

**Hypothesis 1:** Students using a CA for course evaluations will produce a higher response quality compared to a web survey.

**Hypothesis 2:** Students using a CA for course evaluations will perceive higher levels of enjoyment compared to a web survey.

**Hypothesis 3:** Higher perceived levels of enjoyment in course evaluations lead to higher response quality.

**Measurement of Response Quality**

1/3: Normalized Flesh-Reading-Ease (FRE) [5]

1/3: Normalized sentiments

1/3: Normalized self-reported response quality

More information about the measurements can be found in *Measurement and Analysis*

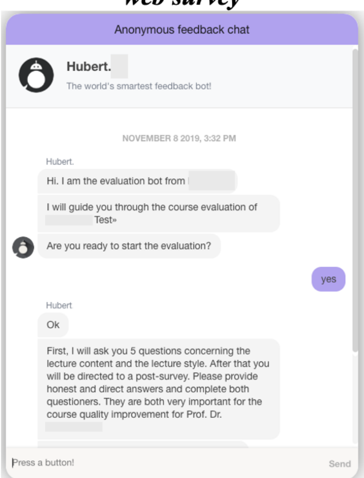
they assist educators with ongoing user feedback to improve their course content and lecture style. However, educators are confronted with certain feedback limitations such as with low acceptance and response rates, only time-related insights and low-quality answers in the open question sections that are hardly applicable for adapting courses to students' expectations [25, 3]. Explanations for these negative effects might be that student responses are affected by survey fatigue [25] or respondents' satisficing behavior [7, 12]. Using evaluations with a static interaction style, such as a web survey, likely leads to low-quality data [12], and this in return makes it difficult for educational institutions to adjust their courses to ever-changing environments. To address these issues, qualitative evaluation methods, such as individual interviews, are used to produce a higher quality of answers and deeper insights [24]. However, these approaches are usually very resource-intensive since lecturers need to address every student individually, which is even more difficult in times of mass lectures such as massive open online courses (MOOCs) [24].

One possible solution to benefit from the advantages of both – qualitative and quantitative – evaluation methods is using conversational agents (CAs). CAs are software programs which communicate with users through natural language interaction interfaces [22, 20]. Compared to traditional quantitative course evaluations, CAs are able to reach students on their everyday devices and build up a human-like interaction with them. The dialogue-based interaction can produce higher levels of enjoyment [4], and therefore might help to overcome the common problems of survey fatigue and satisficing behaviour in course evaluations. Moreover, CAs are able to adapt their answers to students' utterances and can therefore build up a meaningful dialog with the students, almost like a qualitative lecturer-student interview. Backing on social response theory [18, 16, 17], we suggest that this form of human-computer interaction might encourage students to provide a higher quality of answers for lecturers to improve their courses [29]. A recent study by Kim et al. (2019) [12] indicates that a CA can perform part of a human interviewer's role by applying effective communication strategies and, therefore, encourages user enjoyment, which in return leads to high-quality quantitative data. However, literature on the effect of conversational agents on qualitative response data is still scare. Filling this gap, we aim to contribute to the CHI community by investigating *if a CA positively influence the response quality in course evaluation's compared to a traditional web survey.* Therefore, we conducted a field experiment based on social response theory to test weather the different interaction types (conversational vs. static) result in a higher level of response quality of course evaluations. We design an NLP-based CA and deploy it in a field experiment with 127 students in our lecture and compare it with a web survey as a baseline. The findings along with technology acceptance measurements indicate that using CAs for evaluations are resulting in higher levels of response quality and level of enjoyment and are therefore a promising approach to increase the effectiveness of surveys in general. Building on the extensive work in human-CA interaction, our work has implications for the CHI community since we show the influence of the interaction type not only with self-reportings but also with more objective measures (e.g., sentiments and syntactical readability score). Further research is needed to confirm these results and expand our work to similar scenarios.

## Background

Our research is motivated by social response theory. According to social response theory, humans tend to respond socially to agents that display characteristics similar to humans (e.g., to animals or technologies) (Moon, 2000). Behavioral cues and social signals from computers, such as


***web survey***




***conversational agent***

**Figure 1:** Exemplary course evaluation with our CA and our web survey (anonymized)

interacting with others, using natural language or playing social roles, subconsciously trigger responses from humans, no matter how rudimentary those cues or signals are [18, 17]. Following the *"Computers are Social Actors"* (CASA) paradigm, existing research has examined different social cues and their influence on HCI. According to Tung and Deng (2006) [26], students perceive a higher degree of social presence and social attraction in an active-interactivity environment than in a passive-interactivity environment. Also, Schuetzler et al. (2014) [21] showed in their study that a dynamic CA compared to a static interview system is perceived as more engaging. CAs are software programs which are designed to communicate with users through natural language interaction interfaces [22, 20]. In today's world, conversational interfaces, such as Amazon's Alexa, Google's Assistant, Apple's Siri, are ubiquitous, with their popularity steadily growing over the past few years [2, 14]. They are implemented in various areas, such as customer service [31], healthcare [13, 15] or education [10, 30]. While existing research on CAs in education has mainly focused on providing learning support for students [23, 8, 30], Winkler and Söllner (2018) [30] pointed out that CAs might also have potential as an evaluation tool. However, literature that investigates the effect of CAs on course evaluation response data is still scare.

## Method

To test our hypotheses, we conducted a field experiment based on social response theory to test weather the different interaction types (conversational vs. static) result in a higher level of response quality of course evaluations. We designed a field experiment in which students of a large-scale lecture in business innovation were asked to provide feedback on the lecture content and the teaching style. We used a fully randomized between-subject design resulting in the control group (CG) receiving a static interaction type in the form of a web survey and the treatment group (TG) receiving a conversational interaction. The course evaluation questions were exactly the same for both groups, consisting of two quantitative and three qualitative questions following the standard content of the course evaluation of our university. We received 127 valid answers collected from 54 female and 73 male master business students in the first or third semester with an average age of 24,24 years. After randomization 57 students conducted the course evaluation with our conversational agent forming the treatment group, whereas 70 students answered the course evaluation questions with our web survey (CG).

*Design of Course Evaluation Artifacts*

We used two different interfaces: a standard web survey and a CA. The participants of the web survey group (CG) conducted the course evaluation with a web survey tool called *unipark*. We chose this tool since it allowed us to design the survey similarly to the web survey used at our university. The web survey could be completed by the students using either their notebook or a mobile device. The design of the web survey is presented in Figure 1. The quantitative questions were answered with a matrix format to ensure that the same-scaled options were used for multiple items to avoid repeating information. The qualitative items were answered with a plain text input field. For the CA we cooperated with a company specialized on conducting chatbot-based surveys. The cooperation brought several benefits compared to developing our own solution: First, we could rely on proven design experience for questioning bots, which has been applied already in several practical scenarios including course evaluation. Second, the native designed chatbot of the company allowed us to control all design parameters, collect logs of interaction behavior and manipulate the interaction of the CA with the user. Like the web survey, the CA survey could also be conducted using a personal computer or a mobile device. The design of the CA is illustrated in Figure 1.


**Figure 2:** Overview of the experiment phases

*Experiment Procedure*
The experiment consisted of three phases: 1) randomization, 2) course evaluation and 3) posttest (see Figure 2). Randomization and posttest were consistent for both groups. The field experiment started with the lecturer announcing a mid-term course evaluation of the lecture. The students were asked to either type a link into their notebook or scan a QR code with their mobile device. The link led to a web page, which randomly assigned the students to one of the groups. As two students with differently assigned interfaces could be sitting next to each other, they were told that different user interfaces were being tested for improving the design of the course evaluation. The course evaluation was conducted in the middle of the lecture period (after about 50% of the content had been taught). In the course evaluation phase, we asked all participants the same questions: two quantitative and three qualitative questions following the standard content of the course evaluation of our university. The first two questions addressed the perceived benefit of the course ("*I can benefit from the content of the course.*") and the expectation for the lecturing style ("*The lecturer was able to transfer the learning content according to my expectations.*"). Both questions were measured with a 5-point Likert scale (1: strongly disagree to 5: strongly agree, with 3 being a neutral statement). Next, we asked the students three open qualitative questions ("*Which course elements have contributed to your learning success in a positive way?*", *"Which aspects of the course should be changed so that students benefit more from the course?"* and *"Are there any other points you would like to comment on?"*). After the students conducted the course evaluation, they were led to a post-survey, in which we measured different constructs to validate our derived hypotheses.

*Measurement and Analysis*
For measuring the response quality, we used one subjective and two objective measurements: (1) self-reported response quality by the user, and (2) syntactic readability based on the Flesch-readability score [5] as well as the intensity of sentiments in the answers (e.g., [9, 19]). We measured the self-reported response quality by asking participants the following questions: *"The design of the evaluation tool made me think longer about my responses compared to traditional surveys."* and *"I would prefer using a chatbot as a survey tool."* Additionally, we measured the level of enjoyment by asking the following items *"I am satisfied with the evaluation tool."* and *"It is fun to use the evaluation tool."* following Kim et al. 2019 [12]. To measure the syntactic readability of texts, several measures have been used in research [11, 28]. We selected the Flesh-Reading-Ease (FRE) [5] to capture the readability of received responses since this score combines language complexity measurements such as the average sentence lengths and the average syllables per word into one number [5]. The score has been widely used before to determine the readability of a message in computer-mediated communication [27] or for the complexity of CA user responses [6]. Following Flesch (1943) [5], we used the following formula since we received answers in English:

**Flesch Reading Ease** $= 206.835 - (1.015 \cdot asl) - (84.6 \cdot asw)$
***asl****: average sentence length of a response* ***asw****: average syllable per word*

The scores of our answers reach from 0 to 110. The higher the FRE score, the better the readability of the responses. Moreover, we aimed to capture the sentiments of our received responses since a sentiment is a good indicator for an individual taking a position on a certain topic used, e.g., in opinion mining [19]. For example, if a student only an-

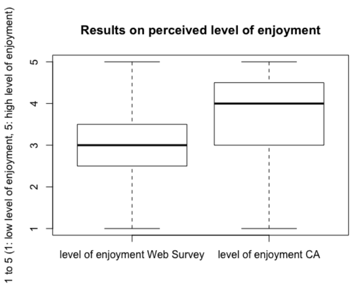



**Figure 3:** Perceived level of enjoyment between web survey and CA

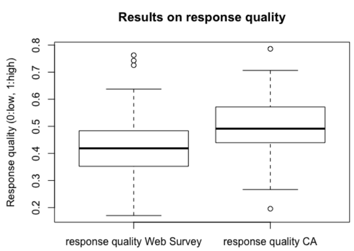



**Figure 4:** Response quality between web survey and CA

swers *"course content in learning unit 2"* no action steps can be derived, since this message has no sentiment (positive or negative notion). Therefore, we used the Naïve-Bayes approach of *TextBlob*, using Python 3.7 to determine the sentiments of each response since it is an easy to use, openly available approach trained on review (evaluation) data. The scores are usually labeled between -1 and 1 according to a "positive", a "negative" and a "neutral" mood (-1 being negative, 0 neutral and 1 positive). However, we multiplied values smaller than 0 with -1 since we did not distinguish between positive or negative sentiments. We believe *"position talking"* sentiments are valuable for the use case of course evaluation, similar to opinion mining [19] or language complexity measurements [9]. Finally, we used a continuous normalized scale from 0 to 1 to measure the sentiments: 0 meaning no sentiment (neutral statement) and 1 meaning high sentiment (no matter if positive or negative). For measuring the FRE and the sentiments, the answers of all three qualitative questions from the course evaluation were combined to one string and analyzed using Python 3.7, utilizing the natural language toolkit (NLTK) [1]. To construct one measurement for data quality, we normalized the construct's self-reported response quality, FRE and sentiments and weighted every measurement with one third to generate one final value to distinguish the responses. In addition, we collected demographic information (age and gender) and asked participants if they had used a CA (e.g., Facebook Messenger Bot) before to control for technology usage between the groups. For data analysis, we used linear regression models and checked their assumptions visually with a test for normality and a test for homoscedasticity. All assumptions are met.

## Results

To investigate, if a CA can improve the response quality and the level of enjoyment of online course evaluations

| | Level of enjoyment | Response quality |
|---|---|---|
| CG*: | M = 3.007, SD = 0.900 | M = 0.429, SD = 0.117 |
| TG**: | M = 3.640, SD = 0.957 | M = 0.499, SD = 0.109 |
| H1 | confirmed | (p = 0.0007282) |
| H2 | confirmed | (p = 0.0001909) |
| H3 | confirmed | (p = 0.00003846) |

**Table 1:** Overview of mean (M) and standard derivation (SD) of the measured constructs *(*n = 70, **n = 57)*

compared to a web survey, we calculated the means and standard derivations (SD) between the control and treatment group (CG and TG) depicted in Table 1. Moreover, we conducted multiple regressions for the corresponding variables for all the hypotheses 1-3. The descriptive statistics are illustrated in Table 1. In total, we received 127 valid answers, 57 for the TG and 70 for the CG. The results of our research model, the r values and the significances are illustrated in Figure 5. To summarize our descriptive findings, we plotted the results on perceived level of enjoyment and response quality between the two interaction types web survey and CA in Figure 3 and Figure 4.

The statistical tests on our results support all of our hypotheses (H1 to H3), meaning that a conversational interaction type significantly influences the perceived level of enjoyment and response quality in online course evaluations. Also, we could confirm our hypothesis H3, meaning that the level of enjoyment in online course evaluations leads to a higher response quality. For all three hypothesis we received p-values smaller than 0.001 as depicted in Table 1. In order to control for potential effects of interfering variables with our sample size and to ensure that the randomization was successful, we compared the difference in the

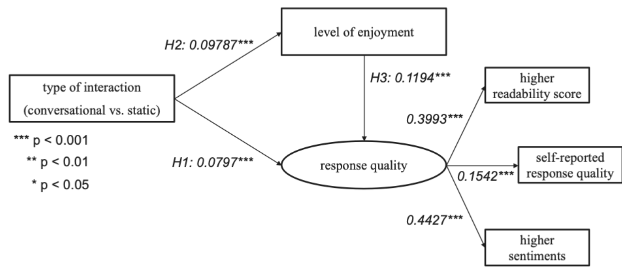



**Figure 5:** Overview of r values and significances of our research model

mean of CA pre-usage. We received p-values larger than 0.05 showing that there was no significant difference between the groups. 95 of the 127 participants had used a CA before (around 75%) across both treatments, where as 32 had not used a CA before (around 25%). Besides, we did not find any significances between the results of the qualitative course evaluations between the treatment groups.

## Contributions, Limitations and Future Work

We found that participants using a CA showed higher levels of enjoyment and were more likely to share high quality feedback. We measured the technology acceptance showing that the intention to use a CA for online course evaluations is higher (mean = 3.52, SD = 0.99) than the intention to use a web survey (mean = 3.39, SD = 0.89). These results are consistent with past studies that investigated beneficial effects of CAs over non-adaptive systems (such as surveys) (e.g., [12]). One reason for the effect might be that conversational interfaces better direct the attention of the user to the question compared to a static web survey [12]. We also show that a CA has the potential to create higher levels of enjoyment in course evaluation. We argue, that this might help to overcome the common challenges of surveys in general, such as survey fatigue [25] or saticsficing behavior [7], and thus leads to better response quality. These results are in line with other studies which show how a CA create a social connection to the user, which increases the perception of being more pleasant and usable [4]. This can also be seen by the qualitative data of our post-survey, where multiple students made positive comments about the interaction with CAs related to the perceived level of enjoyment and interaction: *“It was funnier than the usual web survey”*, *“Chatbot was entertaining to use”*; and related to the perceived usefulness: *“it is very easy to understand and to use”*; related to the response behaviour: *“I think my feedback with this tool is more honest“*; or in general: *“Seemed effortless at first. Made me think i was interacting with a human”*. Our study has several theoretical contributions and practical implications. First, we contribute to HCI research by providing empirical evidence that a conversational interface has positive effects on the answer behavior of respondents. Furthermore, for measuring response quality, we have not only used subjective measurements (self-reportings), but have also applied objective measurements (e.g., number of sentiments and readability). Second, we contribute to the application of CAs in education, suggesting a successful use case to employ a CA with potential benefits for lecturers and educational institutions to better adjust their learning content based on high quality responses and potentially continuous student feedback. Besides, our study faces some limitations. First, we only asked a representative subset of course evaluation questions. Second, it remains open if an ongoing usage of a CA as a course evaluation tool continuously leads to a higher response quality compared to a web survey or if this was only a short time effect. Even if 75% of the participants said they had used a CA before, novelty effects cannot be expelled. Therefore, we call for future work to test the effect of a CA as a course evaluation tool in a longitudinal study.

**Acknowledgements**
We thank *hubert.ai* for supporting our research endeavour by providing the conversational agent.

## REFERENCES


[1] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python*. Vol. 43. 479 pages. DOI:http://dx.doi.org/10.1097/00004770-200204000-00018

[2] eMarketer. 2017. *Alexa , Say What?! Voice-Enabled Speaker Usage to Grow Nearly 130% This Year*. Technical Report. https://www.emarketer.com/Articles/Print.aspx?R=1015812

[3] Malgorzata Erikson, Martin G. Erikson, and Elisabeth Punzi. 2016. Student responses to a reflexive course evaluation. *Reflective Practice* 17, 6 (2016), 663–675. DOI: http://dx.doi.org/10.1080/14623943.2016.1206877

[4] Julia Fink. 2012. Anthropomorphism and human likeness in the design of robots and human-robot interaction. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7621 LNAI (2012), 199–208. DOI:http://dx.doi.org/10.1007/978-3-642-34103-8{_}20

[5] R Flesch. 1943. Marks of readable style; a study in adult education. *Teachers College Contributions to Education* 897 (1943), ix + 69–ix + 69.

[6] Ulrich Gnewuch, Stefan Morana, Marc T.P. Adam, and Alexander Maedche. 2018. Faster is not always better: Understanding the effect of dynamic response delays in human-chatbot interaction. In *26th European Conference on Information Systems, ECIS 2018*. https://aisel.aisnet.org/ecis2018_rp/113

[7] Dirk Heerwegh and Geert Loosveldt. 2008. Face-to-face versus web surveying in a high-internet-coverage population: Differences in response quality. *Public Opinion Quarterly* 72, 5 (2008), 836–846. DOI: http://dx.doi.org/10.1093/poq/nfn045

[8] Sebastian Hobert and Raphael Meyer Von Wolff. 2019. Say Hello to Your New Automated Tutor – A Structured Literature Review on Pedagogical Conversational Agents. (2019).

[9] Aditya Joshi, Abhijit Mishra, Nivvedan Senthamilselvan, and Pushpak Bhattacharyya. 2014. Measuring Sentiment Annotation Complexity of text. In *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, Vol. 2. Association for Computational Linguistics, 36–41. DOI: http://dx.doi.org/10.3115/v1/p14-2007

[10] Alice Kerly, Phil Hall, and Susan Bull. 2007. Bringing chatbots into education: Towards natural language negotiation of open learner models. *Knowledge-Based Systems* 20, 2 (2007), 177–185. DOI: http://dx.doi.org/10.1016/j.knosys.2006.11.014

[11] M. Asif Khawaja, Fang Chen, and Nadine Marcus. 2010. Using language complexity to measure cognitive load for adaptive interaction design. In *International Conference on Intelligent User Interfaces, Proceedings IUI*. 333–336. DOI: http://dx.doi.org/10.1145/1719970.1720024

[12] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing data from chatbot and web surveys effects of platform and conversational style on survey response quality. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–12. DOI: http://dx.doi.org/10.1145/3290605.3300316

[13] Tobias Kowatsch, Marcia Nißen, Chen-hsuan Iris Shih, Dominik Rüegger, Andreas Filler, Florian Künzler, Filipe Barata, Severin Haug, Björn Brogle, Katrin Heldt, Pauline Gindrat, Nathalie Farpour-lambert, and Dagmar Allemand. 2017. Text-based Healthcare Chatbots Supporting Patient and Health Professional Teams : Preliminary Results of a Randomized Controlled Trial on Childhood Obesity. *Persuasive Embodied Agents for Behavior Change (PEACH2017) Workshop* 1, Iva 2017 (2017), 1–10.

[14] Aliane Loureiro Krassmann, Fábio Josende Paz, Clóvis Silveira, Liane Margarida Rockenbach Tarouco, and Magda Bercht. 2018. Conversational Agents in Distance Education: Comparing Mood States with Students' Perception. *Creative Education* 09, 11 (2018), 1726–1742. DOI: http://dx.doi.org/10.4236/ce.2018.911126

[15] Sven Laumer, Christian Maier, and Fabian Tobias Gubler. 2019. Chatbot Acceptance in Healthcare: Explaining User Adoption of Conversational Agents for Disease Diagnosis. *Twenty-Seventh European Conference on Information Systems (ECIS2019), Stockholm-Uppsala, Sweden* (2019), 0–18. https://aisel.aisnet.org/ecis2019_rp/88

[16] Youngme Moon. 2000. Intimate Exchanges: Using Computers to Elicit Self-Disclosure From Consumers. *Journal of Consumer Research* 26, 4 (3 2000), 323–339. DOI:http://dx.doi.org/10.1086/209566

[17] Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues* 56, 1 (1 2000), 81–103. DOI: http://dx.doi.org/10.1111/0022-4537.00153

[18] Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems celebrating interdependence - CHI '94*. ACM Press, New York, New York, USA, 72–78. DOI:http://dx.doi.org/10.1145/191666.191703

[19] Bo Pang and Lillian Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2, 12 (2008), 1–135. http://www.cs.cornell.edu/home/llee/omsa/omsa.pdf

[20] Victoria L. Rubin, Yimin Chen, and Lynne Marie Thorimbert. 2010. Artificially intelligent conversational agents in libraries. *Library Hi Tech* 28, 4 (2010), 496–522. DOI: http://dx.doi.org/10.1108/07378831011096196

[21] Ryan M. Schuetzler, G. Mark Grimes, Justin Scott Giboney, and Joseph Buckman. 2014. Facilitating natural conversational agent interactions: Lessons from a deception experiment. *35th International Conference on Information Systems "Building a Better World Through Information Systems", ICIS 2014* (2014).

[22] Bayan Abu Shawar and Eric Steven Atwell. 2005. Using corpora in machine-learning chatbot systems. *International Journal of Corpus Linguistics* 10, 4 (2005), 489–516. DOI: http://dx.doi.org/10.1075/ijcl.10.4.06sha

[23] Donggil Song, Eun Young Oh, and Marilyn Rice. 2017. Interacting with a conversational agent system for educational purposes in online courses. *Proceedings - 2017 10th International Conference on Human System Interactions, HSI 2017* (2017), 78–82. DOI: http://dx.doi.org/10.1109/HSI.2017.8005002

[24] Carly Steyn, Clint Davies, and Adeel Sambo. 2019. Eliciting student feedback for course development: the application of a qualitative course evaluation tool among business research students. *Assessment and Evaluation in Higher Education* 44, 1 (2019), 11–24. DOI: http://dx.doi.org/10.1080/02602938.2018.1466266

[25] Beatrice Tucker, Sue Jones, and Leon Straker. 2008. Online student evaluation improves Course Experience Questionnaire results in a physiotherapy program. *Higher Education Research and Development* 27, 3 (2008), 281–296. DOI: http://dx.doi.org/10.1080/07294360802259067

[26] Fang Wu Tung and Yi Shin Deng. 2006. Designing social presence in e-learning environments: Testing the effect of interactivity on children. *Interactive Learning Environments* 14, 3 (2006), 251–264. DOI: http://dx.doi.org/10.1080/10494820600924750

[27] Joseph B Walther. 2007. Selective self-presentation in computer-mediated communication: Hyperpersonal dimensions of technology, language, and cognition. *Computers in Human Behavior* 23 (2007), 2538–2557. DOI: http://dx.doi.org/10.1016/j.chb.2006.05.002

[28] Thiemo Wambsganss and Hansjörg Fromm. 2019. Mining User-Generated Repair Instructions from Automotive Web Communities. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, Vol. 6. Hawaii, 1184–1193. DOI: http://dx.doi.org/10.24251/hicss.2019.144

[29] Thiemo Wambsganss, Rainer Winkler, Pascale Schmid, and Matthias Söllner. 2020. Designing a Conversational Agent as a Formative Course Evaluation Tool. In *15th International Conference on Wirtschaftsinformatik*. Potsdam, Germany.

[30] R. Winkler and M. Söllner. 2018. Unleashing the Potential of Chatbots in Education : A State-Of-The-Art Analysis . In : Academy of Management. *Meeting, Annual Chicago, A O M* (2018). https://www.alexandria.unisg.ch/254848/1/JML_699.pdf

[31] Anbang Xu, Zhe Liu, Yufan Guo, Vibha Sinha, and Rama Akkiraju. 2017. A new chatbot for customer service on social media. *Conference on Human Factors in Computing Systems - Proceedings* 2017-May (2017), 3506–3510. DOI: http://dx.doi.org/10.1145/3025453.3025496