Simulated speaking environments for language learning: insights from three cases

Reference:
Sydorenko Tetyana, Smits Tom, Evanini Keelan, Ramanarayanan Vikram.- Simulated speaking environments for language learning: insights from three cases
Computer assisted language learning - ISSN 0958-8221 - 32:1-2(2019), p. 17-48
Full text (Publisher's DOI): https://doi.org/10.1080/09588221.2018.1466811
To cite this reference: https://hdl.handle.net/10067/1510040151162165141

Tetyana Sydorenko[a]*, Tom F.H. Smits[b], Keelan Evanini[c] and Vikram Ramanarayanan[d,e]

[a]*Department of Applied Linguistics, Portland State University, Portland, Oregon, USA;*

[b]*Antwerp School of Education, University of Antwerp, Antwerp, Belgium;*

[c]*Educational Testing Service, Princeton, New Jersey, USA;*

[d]*Educational Testing Service, San Francisco, California, USA;*

[e]*University of California, San Francisco, California, USA;*

*Corresponding author, tsydorenko@pdx.edu

Tetyana Sydorenko (Ph.D., Michigan State University), developer of *SimCon*, is an Assistant Professor of Applied Linguistics at Portland State University. Her research interests include L2 pragmatics, computer-assisted language learning, psycholinguistic processes in SLA, and assessment. She is currently investigating the use of adaptive computer-simulated conversations in the teaching of L2 pragmatics.

tsydorenko@pdx.edu

Tom F.H. Smits

Tom Smits, co-developer of *Papotons!*, is an Associate Professor (Senior Lecturer) for Teaching English and German as a Foreign Language at Antwerp University's Antwerp School of Education. His research involves variation and diversity in the fields of pedagogy (e.g. differentiation) and linguistics (e.g. language variation and change), and covers educational issues (e.g. CALL, EAP) in Belgium, South Africa and the DR Congo.

tom.smits@uantwerpen.be

Keelan Evanini

Keelan Evanini (Ph.D., University of Pennsylvania), co-developer of *HALEF* tasks, is a Research Director at Educational Testing Service. His research interests include automated assessment of non-native spoken English for large-scale assessments, automated feedback in computer assisted language learning applications, and spoken dialog systems. He leads the research team that develops SpeechRater, the ETS capability for automated spoken language assessment.

kevanini@ets.org

Vikram Ramanarayanan

Vikram Ramanarayanan (Ph.D, University of Southern California), co-developer of *HALEF* tasks, is a Research Scientist at Educational Testing Service's R&D division in San Francisco and also holds an Assistant Adjunct Professor appointment in the Department of Otolayngology -- Head and Neck Surgery at the University of California, San Francisco. Vikram's research interests lie in applying scientific knowledge to interdisciplinary engineering problems in speech, language and vision and in turn using engineering approaches to drive scientific understanding. His work on speech science and dialog technology has won 2 Best Paper Awards, an Editor's Choice Award and an ETS Presidential Award.

vramanarayanan@ets.org

Simulated Speaking Environments for Language Learning: Insights from Three Cases

Recent CALL technology reviews cover a plethora of technologies available to language learners to improve a variety of skills, including speaking (cf. Bibauw, François, & Desmet, 2015; Ellis & Bogart, 2007; Eskenazi, 2009; Godwin-Jones, 2009; Golonka, Bowles, Frank, Richardson, & Freynik, 2014). However, few technology-enhanced self-access tools are available for pragmatic development, especially in oral modality. Recognizing the benefits of structured practice for second language development, we demonstrate how such practice can be incorporated into three recently developed simulated speaking environments that vary on the targeted L2 (French, English), domain of use (academic or everyday interaction), emphasis on higher-order and/or lower-order skills, and accommodation of multiple L2 varieties. In the spirit of finding synergies and learning from each other's experiences in specific local contexts, we address the following research questions: (1) How does the local context, researcher and learner goals, and technological possibilities influence the design of each computer application? (2) Based on the examination of the three programs, what can we learn in view of redesign options and suggest to future developers of such programs?

Keywords: intelligent computer-assisted language learning; spoken dialog systems; pragmatic competence; oral practice; simulations

Introduction

Although there are a variety of schools of thought on how second language acquisition (SLA) takes place, in line with Skill Acquisition Theory, we subscribe to the idea that some structured oral practice should be part of language learning (e.g. DeKeyser, 2007, 2010; Lyster & Sato, 2013). This notion also extends to the development of pragmatically appropriate communication - the focus of this paper.

Oral practice can be promoted in a number of ways. Researchers have examined oral storytelling (Kim, 2014), role-plays (Yen, Hou, & Chang, 2015), video production (Sun & Yang, 2015), and telecollaborative exchanges via Skype (Barron & Black, 2015), among others. However, many of these tasks are either not highly structured or do not provide practical opportunities for intrinsic feedback[1] (especially with large numbers of students) nor the modelling needed to move language development forward. Drawing on the psycholinguistic research base, DeKeyser (2010) explains that although declarative knowledge can be developed through other means, like processing of input, in most cases “extensive practice is needed to proceduralize the form-meaning mapping for production and/or comprehension” (p. 158). Additionally, “before students can practice usefully, there needs to be knowledge to be practiced” (p. 161), and one of several ways of doing so is through modelling in communicative interaction. Feedback during contextualized oral practice is likewise necessary. As Lyster and Sato (2013) put it: “repeated practice with feedback at propitious moments promotes the acceleration of *meaningful learning* rather than the acquisition of *mechanical skill* (Anderson, Greeno, Kline & Neves 1981: 206) and thereby contributes to automatization” (p. 72). The value of practice, feedback, and modelling is likewise emphasized in Laurillard’s (2012) Conversational Framework for learning in any area of educational inquiry (e.g. engineering, programming, performing arts). While less structured tasks, like oral storytelling or collaborative work, contribute to language learning by increasing learners’ self-confidence, autonomy, and oral proficiency (Healey, 2007; Kim, 2014) and by helping them reflect on language via interlocutor feedback (Lantolf, Thorne, & Poehner, 2015; Swain & Watanabe, 2013), we follow DeKeyser’s (2010) argument that a portion of language learning should additionally consist of modelled structured practice, planning, preparation, and repetition.

Although some structured oral practice can take place in the classroom, many educational contexts lack sufficient classroom opportunities (e.g. Grobler & Smits, 2017; Sydorenko, Daurio, & Thorne, 2018). Learners in such educational contexts may thus benefit from self-access learning in online and technology-mediated environments. Using technology, learners can maximize opportunities to practice in safe environments (Castellano, Mynard, & Rubesch, 2011; Stracke, 2007) and enhance their autonomous learning skills (Kessler & Bikowski, 2010).

Bibauw, François, and Desmet (2015), Ellis and Bogart (2007), Eskenazi (2009), Godwin-Jones (2009), and Golonka, Bowles, Frank, Richardson, & Freynik (2014) are recent reviews of technologies available to students to improve a variety of skills, including speaking. However, few technology-enhanced self-access tools are available for pragmatic development, especially in oral modality (Taguchi & Sykes, 2013). With this in mind, as well as recognizing the benefits of structured and modelled practice, we compare three recently developed simulated speaking environments and provide ideas for their continued development and improvement. In the spirit of finding synergies and learning from each other's experiences, similar to O'Dowd and Ware's (2009) study comparing telecollaborative exchanges in different contexts, we show parallels and differences between three learning environments and their applicability to specific situations.

Literature Review

Simulated Speaking Environments

Due to recent advancements in technology, the use of simulated speaking environments has been increasing. Spoken dialog systems (SDSs), as one common example of simulated speaking environments, allow learners to have interactive spoken dialogs with the computer system. In many SDSs, learners interact with "embodied conversational agents" – animated characters that are "endowed with conversational capabilities primarily through speech output generation (either synthesized or recorded speech), speech recognition software, and natural language processing" (Morton, Gunson, &

Jack, 2012, p. 2). The agents can produce not only verbal output, but also paralinguistic cues such as frowning, smiling, hand-waving, and walking away. Other SDSs, however, make use of a static image of the automated interlocutor or have no visual agent at all. Due to the limitations of automatic speech recognition (ASR) technology and natural language processing (NLP) algorithms, SDSs generally target restricted conversational contexts so that possible learner utterances are better predicted thus facilitating speech recognition and feedback accuracy (Eskenazi, 2009; Morton et al., 2012).

Some existing SDS-based language learning applications include DEAL (Wik & Hjalmarsson, 2009), DISCO (van Doremalen, Boves, Colpaert, Cucchiarini, & Strik, 2016), ISLAND (McGraw & Seneff, 2007), SCILL (Seneff, Wang, & Zhang, 2004), and SPELL (Morton & Jack, 2005, 2010) (see Bibauw et al., 2015, and Eskenazi, 2009, for detailed descriptions of these and other programs). By focusing on SPELL, Morton et al. (2012) illustrate how such programs generally work. For example, "in the About Train Times scenario, the virtual character asks the learner some questions about the departure and arrival times of trains in Great Britain. To the side of the character on the screen is a timetable depicting the times" (Morton et al., 2012, p. 4). Sometimes there are text help menus to assist learners in overcoming potential difficulties. When learners' utterances are ungrammatical, the virtual character offers implicit spoken feedback, often via recasts. The use of ASR and NLP technologies allows learners to have interactive oral practice with, and receive feedback from, conversational agents.

Morton et al. (2012) found that learners were pleased with the SPELL program and enjoyed using it, although the ASR component did not work perfectly. Continuing challenges include increasing the accuracy of ASR with non-native accented speech and collecting an adequate number of representative learner responses for cataloguing the potential grammatical errors that could inform the feedback provided (Morton et al., 2012). Another challenge, stated in Bibauw, François, and Desmet's (2016) review, is that such systems are either at the prototype stage or they are fully developed but are not available to language learners at secondary or postsecondary institutions (e.g. Tactical Iraqi™, as described by Johnson, 2010).

One of the three systems examined in this paper, *HALEF*, is an SDS. To counter the challenge of ASR errors with certain types of speech input in SDS systems, *HALEF* employs an iterative design and testing approach with a custom ASR component. The two other systems, *Papotons!* and *SimCon*, address this challenge by not including ASR and NLP components, yet still allowing for spoken dialogic interaction to take place.

Pragmatics and Oral Simulated Practice

As stated earlier, the purpose of this paper is to examine specifically the development of pragmatically appropriate communication. Pragmatics can be described in a number of ways, but one commonly used definition is that by Crystal (1997): "Pragmatics is the study of language from the point of view of users, especially of the choices they make, the constraints they encounter in using language in social interaction and the effects their use of language has on other participants in the act of communication" (p. 301). Pragmatics encompasses various aspects of language in use, including production and comprehension of speech acts, pragmatic routines, implicatures, register, politeness, and more (Timpe-Laughlin, Wain, & Schmidgall, 2015). In this paper, the focus is on production of speech acts and pragmatic routines, like making introductions, requests, and ordering food. Scholars agree that provision of authentic input, awareness-raising, communicative practice, and feedback are the necessary components of pragmatic instruction in a second language (L2) (cf. Cohen, 2005; Martínez-Flor & Usó-Juan, 2006); however, from an instructional technology point of view, communicative practice is most challenging (Taguchi & Sykes, 2013).

Early developments of SDS-based systems for pragmatics and speaking practice began with Subarashii (Bernstein, Najmi, & Ehsani, 1999) (for Japanese as L2). Tactical Iraqi™ (Johnson, 2010) is a more recent example of a comprehensive system developed for training soldiers deployed in foreign countries. The latest example of an SDS that includes oral practice of pragmatics is Virtual Reykjavik (Bédi, Arnbjörnsdóttir, & Vilhjálmsson, 2017). Incorporating elements of virtual reality, this application allows learners to practice Icelandic by asking conversational agents for directions in the city of Reykjavik. Bédi's et al. (2017) small scale study of beginning and intermediate-level learner attitudes produced results that are similar to those from other SDS systems: there were some errors in speech recognition, which frustrated some of the learners; many (though not all) learners found the program exciting. Thus, if ASR technology continues to improve, especially with non-native accents, there is promising future for such applications. However, the constraint is that simulated speaking environments must be developed for narrow domains and thus existing programs are not applicable for every learning situation.

In this paper, we examine three simulated speaking environments that incorporated the practice of pragmatics to various degrees. In *Papotons!*, pragmatic competence was a secondary goal within the larger goal of oral communicative competence. The speech acts of greetings, goodbyes, introductions, as well as small talk and formal form of address were the aspects of pragmatic knowledge that were targeted. In comparison, pragmatic competence was the primary goal in *SimCon*, while both pragmatic and oral communicative competence were the primary goals in *HALEF*.

Research Questions

In the rest of the paper, we illustrate and evaluate how both technological challenges and affordances can be tailored to local learning contexts. In designing and developing the three simulated speaking environments (*Papotons!*, *SimCon*, and *HALEF*) as well as in analyzing them for this article, we were guided by the Educational Engineering instructional design model (Colpaert, 2006; 2014), which emphasizes the need for context sensitivity in the search for the most appropriate learning environment. Governed by the Educational Engineering model, we (1) designed each of our applications on the basis of prior research findings regarding the practice of speaking skills and pragmatics, (2) took into account the needs of learners in the local context, (3) considered technological affordances, and (4) included the learning environments' iterative refinement on the basis of our findings. In light of this model, the study addresses the following research questions.

1. How do the local context, researcher and learner goals, and technological possibilities influence the design of each computer application?
2. Based on the examination of the three programs, what can we learn in view of redesign options and suggest to future developers of such programs?

Method

In this section, we describe the design and evaluation of each program. The research questions are then addressed explicitly in the Results and Discussion section.

Papotons!

Description.

The program was developed to foster the interactive linguistic competence of students of French as a foreign language at North-West University (NWU), South Africa. This objective goes beyond grammar or vocabulary exercises but entails, with the principles of task-based learning in mind, engaging learners in encoding and decoding messages in culturally (and thus pragmatically) appropriate ways, and having them deal with communicative and/or information gaps, which in turn implies drawing on their own resources in order to complete the task. The approach is geared to providing meaningful opportunities for individualised learning and feedback through technology (*Papotons!* software package + learning management system (LMS)) in a practice environment blending computer-mediated with face-to-face teaching and learning (see Appendix A). At the core of the program's conceptual design lies the Conversational Framework (CF; Laurillard, 2012), a model that is grounded in theories on what is involved in learning from a teacher and the external environment, and that differentiates between learners' theoretical understanding (e.g. knowledge of grammar) and their practice (e.g. speaking skills). Theory and practice are connected by an iterative relationship (learning cycle): being dissatisfied with their performance may cause learners to reflect upon and modulate their conceptual understanding, which will result in improved performance. Of course, less experienced learners will need help from teachers and that is why, with the CF, teachers have several options to choose from to generate varied and rich feedback loops to motivate and drive learning cycles. Two of these are modelling cycles, representing teachers' as well as fellow students' roles in supporting learning through meaningful task feedback (from teachers, the task itself or through collaboration with peers).

Technology, which according to the CF provides a specific means to motivate and enable learning, consists in the case of *Papotons!* of an interface (see Figure 1) that allows learners to (1) engage in a simulated conversation, (2) benefit from modelling input, and (3) generate an audio file on which they receive audio feedback by the teacher. Peer modelling is provided through a follow-up assignment that has learners practice the same topic by producing a video together.

*** Figure 1 here ***

With *Papotons!*, the students record their individual contributions to a basic simulated conversation (e.g. introducing oneself) after listening to each of the pre-recorded questions as many times as they wish. They can also listen to their answers and re-record them if necessary. If they are unsure about the meaning of the question or the format of the response, a model dialog can be consulted to obtain intrinsic feedback. At the end of the simulated dialog, an audio-recording of an individual modelling the successful completion of the activity is provided before the students are required to rate their own performance in four areas (see Figure 2).

*** Figure 2 here ***

A software-generated sound file is made available to the teacher (through the university's LMS), who can subsequently produce an integrated sound file consisting of the initial question, the student's response and the teacher's feedback. By asking students to do a short quiz on the individual feedback, the teacher can ensure that they listen to the feedback at least once, and their reactions help the teacher to adapt and improve their feedback.

Finally, students are required to work in pairs on the production of a short video clip (with their mobile phones) on the same topic as the practice dialog. The video clips are made available on the LMS where other students give advice and/or feedback and/or commentary on these productions (see Laurillard, 2012, p. 89–90).

The simulated conversation task involves responsive oral output (Brown, 1994), that is, short replies to questions that serve as a "preliminary stepping stone" towards real communication (Pino-James, 2013, p. 40). The program thus gradually prepares students for a face-to-face communication situation. The opportunity to listen to a model of the simulated conversation assignment allows the learner to "see the result of their actions in comparison with the intended model [, which] is one of the most powerful ways of learning, and enables the learner to 'learn without being taught'" (Laurillard, 2012, p. 90). The activity of rating personal performance after completing the simulated conversation adheres to the recommendations of Eisenmann and Summer (2012, p. 418) to provide students "with means of assessing their own performance."

Throughout the process, students have the opportunity to make use of different types of support: technical support (contact sessions in language laboratory, e-mail contact with software developers), facilitator (content support), and teacher support.[2] Without such support, learners tend to use fewer or inefficient learning strategies, motivation levels tend to be low and dropout rates high (Reinders & White, 2010). This support facility notwithstanding, one of the greatest affordances of the program is an increase in self-regulation – thanks to feedback and interactivity (e.g. Figura & Jarvis, 2007) – which has been proven to correlate with self-efficacy: self-regulation is contingent on positive self-efficacy beliefs (Winne, 2005). Both can be seen as crucial factors of success across a variety of learning contexts (Shea & Bidjerano, 2010).

Evaluation.

With a view to identifying the contextual issues that are fundamental to the (design of the) *Papotons!* practice environment, data have been collected on the users' levels of self-directed learning, the degree of their foreign language anxiety, their technology acceptance, and computer self-efficacy. Self-directed learning and computer self-efficacy will be discussed here (for additional details, see Grobler & Smits, 2017). After a process of validation and cultural adaptation, Cheng et al.'s (2010) self-directed learning inventory (SDLI) served to gauge the degree of teacher support necessary when students participate in out-of-class practice activities. The SDLI consists of four factors: learning motivation (of paramount importance to effective SDL), planning and implementation (learning methods, strategies and skills for SDL), self-monitoring (selfreliance and self-regulation), and inter-personal communication (learner interaction and ability to express oneself). The questionnaire was distributed twice, each time a week before a practice cycle was launched. It uses a scale of 1 to 5, with 1 being ‘strongly disagree’ and 5 ‘strongly agree’ with 20 statements on how students approach their studies. 48 students were approached, but only 21 completed both questionnaires (also see Appendix A).

*** Table 1 here ***

Learning motivation and self-monitoring scored the highest (even higher if the students who only took part in the first round of data collection are included) and continue to do so in the course of the semester. With regard to planning and implementation and interpersonal communication, the results indicate that students are more neutral (i.e. less confident) as to their self-perceived ability to plan and implement SDL as well as to express themselves and interact. The descriptive statistics indicate that all four aspects of the inventory resulted in a decrease in the second round; however, none of these differences were statistically significant ($p < .05$, see Table 1). The effect sizes indicate that the largest decrease was in motivation (large effect, following *Cohen’s d* guidelines[3]). The decrease in motivation may be due to a bias caused by falling participant numbers or by the fact that the novelty had worn off by the second round. We return to this point in the Results and Discussion section. Students' self-efficacy was high (an average of 4.5 out of 6 on Cassidy and Eachus’ (2002) computer user self-efficacy questionnaire), indicating that students see themselves as highly self-efficient with regard to computer use. This information contributes to iterative design of the simulated speaking environment.

SimCon

Description.

The second program, *SimCon*, was designed for the practice of extended pragmatic routines and appears to be best-suited for intermediate to advanced learners (rationale provided in the Results and Discussion section). While the program is not language-specific, development and evaluation efforts thus far have focused on adult English language learners in the U.S.

In this program, learners complete simulated conversations on a particular topic within a restricted domain, such as making a certain request to an instructor. Each simulation begins with a scenario description, as depicted in Figure 3.

*** Figure 3 Here ***

The first step for learners is to watch the video that initiates the simulation. This scenario begins with a video of an instructor who is typing on the computer in his office. Next, learners record their spoken response to the video they have just watched. For example, one of the participants said “Hi Dan. Can I come inside?” After this, learners are taken to the next screen, where they can play back the response they have just recorded and select the option that best matches the action they have just completed (see Figure 4). In the case of this particular learner, it should be “You greet the instructor and ask if you can come in.” Note that response options are possible actions, not possible verbatim statements; according to survey results, this constitutes feedback to learners on possible actions after each video (see Sydorenko, et al., 2018).

*** Figure 4 Here ***

The program then takes learners to the video that corresponds with their selection. For the selection depicted in Figure 4, the corresponding video is of the instructor who says “Yeah, definitely. What can I do for you?” Connections between response options chosen and the next videos that appear constitute pragmatic feedback in the form of consequences to learners' actions. After this, learners continue with a branching conversation in the same stepwise fashion as described above, alternating between watching videos, responding orally, and selecting from a list of possible options.

Theoretically, learning in *SimCon* can take place via the following mechanisms. First, users can practice their responses as students to an instructor in a particular situation an unlimited number of times: oral rehearsal alone, without any feedback, can lead to improvements in fluency, accuracy, and complexity (see Ellis, 2009). Second, *SimCon* tasks enable the learners to reverse roles and respond as instructors to student videos; in doing so, participants can witness authentic interactions involving expert speakers in a similar situation and can incorporate such model input into their own production (Sydorenko, 2015; Sydorenko & Tuason, 2016). The postulated learning mechanism at play is that learners notice gaps in their production as compared to that of more expert interlocutors (Gass & Mackey, 2015). Third, feedback in the form of positive or negative consequences from following videos and ideas for possible next actions may draw learners' attention to additional opportunities for learning (see Sydorenko et. al., 2018, for more details).

Evaluation.

14 intermediate-level and 18 advanced learners of English (out of twenty learners approached in each proficiency group) volunteered to interact with the simulations and provide their opinions about the program. However, due to technical difficulties in the advanced group, the data from only 11 learners were gathered. Three kinds of data were collected: users' attitudes towards *SimCon* (intermediate and advanced students) and learning outcomes and users' perceptions of learning outcomes (advanced students only). (For additional details, see Sydorenko, 2016; Sydorenko, et al., 2018, and

Appendix B). To assess attitudes, learners completed a survey about the program's authenticity[4] (questions 1 and 2) and ease of use (question 3). Participants were asked to answer each question with "yes," "mostly yes," or "no," as well as to provide a reason for their response. There was also space to provide additional comments about the program.

Descriptive statistics in Table 2 indicate that participants viewed their interactions with the program as mostly resembling real life. However, there were group differences: 14% and 22% of intermediate-level English as a Second Language (ESL) students answered "No" to questions 1 and 2, respectively, compared to 0 "No" responses to these questions from advanced learners. Judging by the amount of "No" responses to question 3, the program was easier to use for advanced learners than for intermediate learners.

*** Table 2 here ***

Learners' qualitative responses that further speak to the authenticity of the program addressed the naturalness of background in the videos (e.g. interior design of the instructor's office), the realistic appearance of the instructor, and nervousness students felt when completing the simulations. Learners' additional comments indicate that while they liked the simulations used, they would welcome more varied topics and more possible response options.

In terms of learning outcomes, only the advanced learners (same as those represented in Table 2) were studied. Learners completed six simulations: simulations 1 and 2 served as a pretest; simulations 3 and 4 were on the same scenarios as 1 and 2 (requesting a letter of recommendation in different circumstances – see Table 3), except learners played instructor roles while getting models of student videos; simulations 5 and 6 were the same as 1 and 2 and thus served as a posttest. By comparing the differences in learners' oral responses to pretest and posttest simulations, it was found that (1) learners produced more model than personal changes and (2) there were more content than form changes (see Table 3). We defined model changes as those that approximated form and content from models, while personal changes were not related to models. Form, like Long's (1991) definition of focus on form, stands for linguistic form (vocabulary, grammar, pronunciation). Content includes changes in semantic content, often in a form of sociopragmatic strategies (see Appendix B for more details). The results indicate that learners made mostly appropriate changes in vocabulary, pragmatic strategies, and cultural notions (examples are provided in Appendix B). Learners' perceptions of learning outcomes were in line with the findings from oral production data in that learners also felt that they gained situation-specific appropriate vocabulary, pragmatic strategies, and cultural concepts. Some learners also stated that they gained an increased level of confidence for communicating with target language speakers in similar situations. Additional qualitative findings are provided in the Results and Discussion section where the themes of authenticity, modelling, and feedback are addressed. For full details on findings, see Sydorenko, et al. (2018).

*** Table 3 here ***

HALEF

Description.

HALEF[5] (Help Assistant--Language-Enabled and Free) is the final CALL paradigm we discuss in this paper. As in the case of *SimCon*, development and evaluation have concentrated on ESL. *HALEF* is an interactive, open-source SDS with audio and video support that is cloud-based and industry standards-compliant (see Ramanarayanan et al., 2017 for additional details). The *HALEF* infrastructure has been used to develop interactive CALL tasks that process a language learner's speech using ASR and NLP in real-time so that the system's subsequent prompt is determined by the content of the learner's utterance. A wide range of goal-oriented tasks have been explored using *HALEF*, such as responding appropriately to compliments and making requests in a workplace environment (Timpe-Laughlin et al., 2017), interviewing for a job, requesting a refund, and taking customer orders. In this paper, we will focus specifically on a task that was designed to provide learners with an opportunity to practice ordering food in a restaurant, specifically a coffee shop.[6]

In the Coffee Shop *HALEF* task, learners are instructed to purchase a beverage and a food item for their boss. After clicking on a button labelled "Speak with the barista" to initiate the conversation, learners hear the barista ask the following question: "Hi. Welcome to the Coffee Spot. What can I get you today?". They can then place an order for a single item or multiple items, and the system subsequently asks follow-up questions to find out details about the order. For example, if the learner orders a coffee, the system will ask for the size (small, medium, or large), whether it should be iced or hot, etc. If the learner specifies this information in their initial utterance (e.g. "I'd like a large iced coffee with cream, please.") these details are recognized by the system and the associated follow-up questions are bypassed. After asking follow-up questions about each item that the learner orders, the barista asks "Can I get you anything else?" If the learner orders another item, the cycle is repeated; if not, the system asks a few final questions, such as whether the order is for here or to go (the targeted appropriate answer is "To go" since the order needs to be brought to the boss), which payment method will be used (cash or credit), etc., and ends the conversation. Figure 5 presents a screenshot of the Coffee Shop task page including the instructions that are presented to the learner, the image of the barista, and a menu of items that can be ordered.

*** Figure 5 here ***

At the end of the conversation, learners are presented with post-task feedback about task completion; specifically, they are told whether the system heard an order for a beverage and an order for a food item (as requested by the boss). While errors in ASR sometimes lead to missed orders, the system generally performs well with high-quality audio input (little background noise and a headset microphone). The system is designed to only respond to on-task utterances and provides a generic response ("I'm sorry, I don't understand. Could you repeat that?") if the learners provide an off-task response; this constraint helps to improve the ASR performance, but limits the range of responses that learners can provide (this trade-off will be discussed further in the Results and Discussion section).

Evaluation.

The data sample considered in this paper consists of 7,345 utterances from 849 different conversations collected from non-native speakers of English interacting with the Coffee Shop task via the Amazon Mechanical Turk platform between June 2016 and February 2017. The English learners in this sample are represented by a total of 52 different first languages (L1s); the 10 most frequent L1s in this sample are shown in Table 4 along with the number of learners for each.

*** Table 4 here ***

As Table 4 shows, a substantial percentage of the participants are from India, which is consistent with the demographics of the overall Amazon Mechanical Turk population. However, many other L1s are represented in this data set as well, providing ample opportunities to study variations across the learner responses due to differences in local context.

In order to evaluate user experience during their dialog interactions, we asked participants to rate their interactions on several metrics (see Table 5).

*** Table 5 here ***

Figure 6 plots the survey results for the coffee shop interactions we analyzed as part of the study reported here. We generally observe that most participants' experiences were very positive, with the largest proportion of participants assigning a rating of 5 for each metric, and more than two-thirds of participants reporting ratings of 4 or 5 for all metrics.

*** Figure 6 here ***

Results and Discussion

Research Question 1: How do the local context, researcher and learner goals, and technological possibilities influence the design of each computer application?

Although all three computer applications address the practice of pragmatics in the oral modality, following Educational Engineering principles, each program was designed differently. In our comparisons of the three programs, in addition to the broad factors of context, researcher and learner goals, and technological possibilities, the following specific differences were observed and are described in detail below: learner proficiency levels, focus on lower order thinking skills (LOTS) or higher-order thinking skills (HOTS), provision and types of modelling and feedback, and use of self-evaluation.

Papotons!

This program was designed for beginning learners of French in large classes who, being in a non-target language environment, had limited oral practice opportunities; for this reason, the application included very simple prompts and responses. In terms of pragmatics, greetings, introductions, good-byes, small talk, and the formal form of address were targeted. Simplicity also typifies the software's design, allowing it to be used in low-resource settings (e.g. off-line, minimally visual). As described above, the goal of the program is to gradually prepare students for a face-to-face communication situation. Thus, the simulations themselves focus on LOTS, such as remember, understand, and apply (i.e. produce a similar conversational response given the model. (In categorizing LOTS and HOTS, we use Krathwohl's (2002) revised version of Bloom's taxonomy.[7]) However, learners are presented with an opportunity to use HOTS, such as evaluate and create, in subsequent tasks of self-evaluation and creation of videos with peers. While HOTS have been recently emphasized, we agree with Colpaert (2016) and Gijsen and Colpaert (2017) that context must be taken into consideration and therefore that emphasis on LOTS, HOTS, or both types of skills can be most appropriate when designing a context sensitive language learning environment.

Because of learners' low proficiency level, both modelling via pre-planned audio within the program and detailed feedback by the teacher were included. Given the lack of L2 contact in the local context, it was determined that modelling and feedback were critical in moving language development forward. To encourage autonomy, learners were also asked to self-evaluate their performance. Modelilng, feedback, and self-evaluation were further reinforced by peer modelling through joint video creation. The design of *Papotons!* software, LMS, and activities around these applications makes it clear that the program was envisioned as an integrated part of curriculum supported by peers and the teacher. On the other hand, *SimCon* and *HALEF* were designed as self-access programs.

SimCon.

In theory, this program could be useful for practicing pragmatic routines by learners of any level as long as the given situation is within a restricted context where responses to each video are maximally predictable. However, given the high-stakes challenging situations that were tested (e.g. asking an instructor for a letter of recommendation due in three days), the naturally fast-paced speech in videos (for authenticity reasons), and comparison of attitudes between intermediate and advanced-level learners, it appears that the given simulations are more appropriate for advanced levels. The advanced learners (Fulbright scholars) were about to begin their Masters and Doctoral studies in the U.S., thus it was deemed appropriate and timely to prepare them for the scenario of requesting a letter of recommendation.

In *SimCon's* letter of recommendation simulation, both HOTS and LOTS can be practiced.LOTS (like remember and understand) were most accessed when greeting the instructor and practicing pronunciation of particular words. Learners also had a chance to practice HOTS (such as analyze) when strategically planning for sequencing the request. For example, when should one say that a letter of recommendation is due in three days: after they have asked for the letter and received instructor's agreement to write it, or before this? Many of our participants experimented with such sequencing.

Since minimal feedback was provided due to limited technological possibilities (see below), modelling was especially critical: native speaking student videorecorded responses provided learners with appropriate pragmalinguistic expressions like "I was wondering if you could," hedged language like *maybe, just,* and *sort of*, and sociopragmatic strategies, such as disarmers ("I know this is really last minute") and imposition minimizers ("I understand if you don't have time") (see Sydorenko et al., 2018, for details). However, although simultaneously receiving models of student behaviour and enacting instructor roles might make the task more exciting, it also makes it more challenging: learners need to not only enact instructor roles, which is difficult for someone who has not been an instructor (see Sydorenko, 2015), but also pay attention to model input from student videos and take mental note of what in that input is different from their responses. While advanced students were able to do this to some degree, several mentioned they would benefit from taking notes during the simulation. This suggests that cognitive load during the performance of these simulations was high. Given numerous appropriate changes learners made, cognitive load may not need to be reduced for advanced learners; however, it remains to be seen whether learners of other profiles (lower level, less autonomous) would benefit from this level of cognitive load.

Feedback provided via the program was implicit: learners received consequences to their actions via subsequent videos, as well as lists of possible next actions. This design choice was made for two reasons. First, since the program does not use NLP and ASR, automatic individualized feedback is not possible, and teacher-provided individualized feedback is not ideal in self-access programs. Designers of CALL programs can often be limited by technological possibilities (Strik, Cornillie, Colpaert, van Doremalen, & Cucchiarini, 2009), and this was the case for feedback in *SimCon*. Second is the elusive nature of appropriate pragmatics feedback: what one target language speaker may find appropriate in a given situation may not be deemed appropriate by another (especially in the simulations used in *SimCon*, as evidenced by our pilot data). Instead, student models to which instructors are likely to respond more favourably were provided. Thus, in terms of pragmatics, the primary affordance of simulations used in *SimCon* is raising learners' awareness of factors to consider when communicating with instructors. As survey data reveal, *SimCon* also contributes to the increase in confidence to communicate. We consider this a valuable outcome as learner anxiety in interacting with target language speakers is a well-documented phenomenon (e.g. Duff, 2007).

HALEF.

Like *SimCon*, *HALEF* was also designed as a self-access program in which both LOTS and HOTS can be practiced, depending on the nature of the communicative task. Because tasks can vary widely, *HALEF* was envisioned to be applicable for learners of various levels. The analysis in this paper focuses on the Coffee Shop task, which involves primarily LOTS since the conversation is transactional in nature (see Evanini et al., 2017 for further analysis of the Coffee Shop task). The learners' utterances in this task are typically short and provide specific information based on the system's questions, such as *What can I get for you?* or *Would you like that coffee hot or iced?* This design approach for an SDS-based conversation is referred to as *system-initiated* since the conversational flow is driven by the system's utterances and leads to more accurate linguistic processing of the learner's utterances. Other design approaches that enable users to interrupt the SDS and to have more control over the conversational flow are possible, but are more difficult to design in a robust manner.

HALEF is designed to process a user's response in real-time and determine the next system prompt based on the content of the user's response; this functionality enables branching spoken dialogs in which the system only provides prompts that are appropriate given the preceding dialog context. For example, if the customer orders a drink in the Coffee Shop task, the system asks questions about the drink order (such as size, hot vs. iced, etc.); alternatively, if the customer orders a food item, the system asks a different set of questions (such as whether it should be toasted or not).

One of the key differences between the Coffee Shop task in *HALEF* and the tasks described in the other two systems is the focus on one or multiple varieties of an L2. Given the status of English as a global language (Crystal, 1997) and a recent call to examine the pragmatics of English as a lingua franca (House, 2009), the designers' goals were to provide learners with an opportunity to speak different varieties of

English. We envisioned a local context of an international company. In a coffee shop located on such imaginary company's campus employees are used to interacting with customers who speak different varieties of English. Thus, one of the main challenges for the designers is to enable the system to accurately process a wide range of potential variations in the learners' responses. An example of this can be found in some of the diverse responses to the following system prompt (the L1 of the learner who provided each sample response is indicated in parentheses after the response):

SDS prompt: *Would you like that for here or to go?*

Responses: *Make it take away.* (Kannada)

Hi, I want some, some coffee to take away. (Spanish)

It's on a go. I I need it as a parcel. (Hindi)

Based on the task designer's initial expectation of the types of responses that learners would provide to this question, the natural language understanding module used keyword-based regular expressions for the two semantic categories: *here* and *stay* (for the HERE category) and *to go*, *carry out*, and *take out* (for the TO_GO category). The sample responses listed above are problematic, since the semantic category clearly should be TO_GO, but learners used variable ways of expressing this meaning that were not initially expected by the system designers. The most effective way for addressing this type of local variation in learner responses is to systematically transcribe and provide semantic annotations for a large number of responses in order to capture as many of the different variants as possible. Then, system designers can use the annotated responses to develop additional key words and phrases for the semantic categories for a rule-based natural language understanding module. In this case, the phrases *take away* and *as a parcel* could be added to the list for the TO_GO semantic category in order to increase coverage. This iterative approach to system design is necessary in order to be able to develop high-performing interactive SDS tasks that can accurately process patterns of responses from learners in different contexts (e.g. different L1s, proficiency levels, etc.).

We will now summarize and synthesize the considerations for the design of each program. As Table 6 illustrates, the three contexts varied: *Papotons!* was designed for a specific population of learners of French and tested on that population; on the other hand, both *SimCon* and *HALEF* programs were designed for a variety of learners of English. It was envisioned that while maintaining the same program structure, tasks in *SimCon* and *HALEF* will need to be varied depending on specific learners' needs. The *SimCon* tasks described here were created for ESL learners (both intermediate and advanced level) who were preparing to begin degree programs in the U.S. and thus wanted to practice communicating with target language speakers in an academic environment. For the *HALEF* task, the goal was to examine how well the program can accommodate a variety of global Englishes. Learner and designer goals of practicing oral communication were the same, but all designers additionally recognized that particularly culturally and pragmatically appropriate communication is key and that learners generally lack opportunities for such practice. Regarding LOTS and HOTS, all designers recognized that both should be practiced, but that their applicability is dependent on the specific simulated task/communicative function. Since *Papotons!* was used to supplement in-class instruction, there was more flexibility in that context (as compared to the other two self-access contexts) for addressing skills that were not easily accessible via tasks in simulated environments. Available resources in each context also determined whether advanced ASR and NLP technologies were used and thus whether and how modelling and feedback were provided (also see the results of research question 2 for future plans regarding feedback).

*** Table 6 about here ***

Research Question 2: Based on the examination of the three programs, what can we learn in view of redesign options and suggest to future developers of such programs?

By comparing the results from the three programs, we observed that changes can be made with regard to motivation, authenticity, modelling, feedback, and self-directed learning.

Information on motivation can play an important role when investigating CALL systems and may explain practice efforts and learning gains, amongst other things (Bodnar, Cucchiarini, Strik, & Van Hout, 2016). As Table 1 indicates, learners' motivation after using *Papotons!* throughout the semester began to decrease. Some possibilities are that the level of task complexity had become too low to motivate students as their language proficiency was growing,or that the novelty and attraction of the program wore off and subsequent simulations were not different enough to keep the motivation high (e.g., Sydorenko, 2011). In *SimCo*n, while the majority of learners stated they enjoyed the program, many suggested topics for other simulations they would like to try. Motivation and topical preferences were not assessed in the case of *HALEF*; however, it is conceivable that some Amazon Mechanical Turk participants were more motivated by the compensation for task completion than by their learning outcomes. Thus, follow up *HALEF* studies will be conducted with participants enrolled in language learning programs and will include an assessment of learners' motivation. In line with Gijsen and Colpaert's (2017) proposal, learners' topical preferences will also be collected; however, it remains to be seen whether more challenging simulations will be accessible to lower-level learners and whether technological restrictions will allow to make simulations on topics suggested by learners robust.

Authenticity is the second aspect of user experience that became apparent. For example, many *SimCon* users commented on the instructor's realistic demeanour and feeling nervous when completing the simulations (a desired outcome, given the high-stakes situation ofthe recommendation letter due in three days).For authenticity reasons, *HALEF* tasks also incorporate visuals, such as the menu in the Coffee Shop task. Given the low-resource environment, the current versions of *Papotons!* does not use visuals to enhance authenticity; however, this is a possible future development. Future research can examine if visuals, and which ones, might make simulated interactions feel more authentic. For example, should these visuals be pictures, videos, or avatars? Does this depend on a particular simulation or the level of the learners?

Next we discuss modelling and feedback, and relations between them. As evidenced by participants' opinions of and learning outcomes from *Papotons!* and *SimCon*, modelling contributed to their language development and was noticed and incorporated into their production. However, *SimCon* participants, although obtaining some ideas from implicit feedback, wished for more individualized feedback. These findings are in line with other researchers (e.g. DeKeyser, 2010; Laurillard, 2012; Lyster & Sato, 2013) who believe in the necessity of modelling, practice, and feedback. Among these important components, automated and individualized feedback is technologically most challenging. For this reason, in *Papotons!*, individualized feedback is provided by the teacher. In *HALEF*, post-task feedback is currently limited to task completion, e.g. whether the learner ordered both a drink and a food item in the Coffee Shop task. In the future, due to advances in ASR and NLP, individualized automatic feedback about language use, such as pragmatic appropriateness, will also be implemented using native speaker responses as a baseline. Although pragmatic appropriateness is not always an objective measure, certain statements, like "I want you to open the door" addressed to a superior, can be seen as clearly inappropriate. Based on the errors learners made, remedial exercises can be also provided within the system (similar to the program design in van Doremalen et al., 2016). In *SimCon*, individualized automatic feedback is not possible. Instead, a generalized feedback option with detailed explanations of cultural and linguistic notions that affect pragmatic appropriateness is now being developed. Subsequently, learners will be asked to compare their responses to generalized feedback and to reflect on changes they could make. The goal of such generalized feedback function is to raise learners' pragmatic awareness rather than to provide them with answers for each situation they might encounter.

Self-directed learning is a final aspect of a simulated speaking environment that we reflect on. The aspect of self-evaluation available in *Papotons!* clearly exemplifies the goal of developing learner autonomy. Such elements could similarly be incorporated in the other two programs examined, although, given that *SimCon* and *HALEF* are designed for self-access, learners would need to have the motivation and self-directed learning goals to engage in self-evaluation However, we would like to take this idea a step further and consider the effect of individual learning differences. For example, in collecting learner feedback on their SDS focused on practicing grammar during spoken interaction, Strik et al. (2009) learned that their participants had varied preferences for corrective feedback. Thus, when resources allow and technological possibilities exist, it would be worthwhile to design and evaluate a program that allows users to choose the way they learn best. One should not forget, however, that autonomy depends on learners' ability to self-direct for practice, critical reflection, and independent action

(Andrade, 2012; Little, 1991), and thus programs that provide learners with choices may not be suitable for all or instructor facilitation in the use of such programs may be necessary.

Conclusion

In this paper, we illustrated how the local context, researcher and learner goals, and technological possibilities affect the design of simulated speaking environments. *SimCon* tasks do not necessarily need to address high-stakes academic situations and target advanced learners; in a context with more resources and more proficient learners, *Papotons!* program itself (rather than post-*Papotons!* in-class work) could focus on both LOTS and HOTS; *HALEF* tasks do not need to account for L2 variation, and in fact many other *HALEF* tasks do not accommodate variation. However, the various programs were designed and used in the ways discussed above because they addressed learners' needs and/or designers' goals in given contexts. Technological limitations likewise played a role in the design. For example, since immediate automatized feedback is not possible in *Papotons!* and *SimCon*, modelling was relied on heavily. While our evaluations of the three programs in varied contexts have allowed us to provide tentative considerations for the design of similar programs for other contexts, we must acknowledge that *Papotons!*, and especially *SimCon*, were tested on very small learner populations and the results are thus not robust – studies with larger learner populations in specific contexts are thus advisable before firm conclusions are drawn. Additionally, motivational aspects of the three programs and particular simulations/topics could be examined in more detail qualitatively. Finally, while all three programs were designed for oral pragmatics practice, only *SimCon* was evaluated for pragmatic learning outcomes. For this reason, the studies with all three programs are ongoing and include the redesign-evaluation-redesign cycles.

We conclude with the words of Colpaert, Cucchiarini, Strik and Oberhofer (2011), “A particular technology can be useful in one LE [learning environment], but useless in another LE. No technology can be evaluated in a discrete, absolute way. No LE can be evaluated independently from its context.”

Acknowledgement

This paper was inspired by the theme of the CALL 2017 conference: CALL in Context. Among many others, Dorothy Chun’s (2017) presentation provided illustrative examples of context effects on the success of a telecollaboration, while Linda Gijsen and Jozef Colpaert (2017) cautioned that among a multitude of available resources, pedagogues should not settle on the most common ones but consider all options given a particular context.

References

Andrade, M. S. (2012). Self-regulated learning activities: Supporting success in online courses. In J. S. Moore & A. D. Benson (Eds.), *International perspectives of distance learning in higher education* (pp.111–132). Rijeka, Croatia: InTech. Retrieved from http://dx.doi.org/10.5772/33745

Barron, A., & Black, E. (2015). Constructing small talk in learner-native speaker voice-based telecollaboration: A focus on topic management and backchanneling. *System, 48*, 112-128. https://doi.org/10.1016/j.system.2014.09.009

Bédi, B., Arnbjörnsdóttir, B., &, Vilhjálmsson, H. H. (2017, July). *Learners' expectations and experiences in Virtual Reykjavik*. In Proceedings of Computer Assisted Language Learning (CALL) Conference, Berkeley, CA. Retrieved from http://call2017.language.berkeley.edu/wp-content/uploads/2017/07/CALL2017_proceedings.pdf

Bernstein, J., Najmi, A., & Ehsani, F. (1999). Subarashii: Encounters in Japanese spoken language education. *CALICO Journal*, *16*, 361-384.

Bibauw, S., François, T., & Desmet, P. (2015). Dialog-based CALL: an overview of existing research. In F. Helm, L. Bradley, M. Guarda, & S. Thouësny (Eds), *Critical CALL – Proceedings of the 2015 EUROCALL Conference, Padova, Italy* (pp. 57-64). Dublin: Research-publishing.net. http://dx.doi.org/10.14705/rpnet.2015.000310

Bibauw, S., François, T., & Desmet, P. (2016). *Instructional design and natural language processing in dialog-based CALL.* Paper presented at CALICO, East Lansing, MI.

Blum-Kulka, S., House, J., & Kasper, G. (Eds.). (1989). *Cross-cultural pragmatics: Requests and apologies*. Norwood, NJ: Ablex.

Bodnar, S., Cucchiarini, C., Strik, H. & Van Hout, R. (2016). Evaluation of the motivational impact of CALL systems: current practices and future directions. *Computer Assisted Language Learning, 29*, 186-212.

Brown, H. (1994). *Teaching by principles. An interactive approach to language pedagogy*. Englewood Cliffs: Prentice Hall.

Cassidy, S. & Eachus, P. (2002). Developing the computer user self-efficacy (CUSE) scale: investigating the relationship between computer self-efficacy, gender and experience with computers. *Journal of Educational Computing Research, 26*(2), 133-153.

Castellano, J., Mynard, J., & Rubesch, T. (2011). Student technology use in a self-access center. *Language Learning & Technology, 15*(3), 12–27. Retrieved from http://llt.msu.edu/issues/october2011/actionresearch.pdf

Cheng, S., Kuo, C., Lin, K., & Lee-hsieh, J. (2010). Development and preliminary testing of a self-rating instrument to measure self-directed learning ability of nursing students. *International Journal of Nursing Studies, 47*, 1152–1158. doi:10.1016/j.ijnurstu.2010.02.002

Chun, D. (2017, July). *Contextual challenges of Telecollaboration 2.0*. Paper presented at the Computer Assisted Language Learning (CALL) Conference, Berkeley, CA.

Colpaert, J. (2006). Pedagogy-driven Design for Online Language Teaching and Learning. *CALICO Journal*, *23*(3), 477–497.

Colpaert, J. (2014). *Educational Engineering and Distributed Design*. A Research Report. EuroCall 2014. Groningen. Retrieved from http://www.jozefcolpaert.net/EE.pdf

Colpaert, J. (2016). Big content in an Educational Engineering Approach. *Journal of Technology and Chinese Language Teaching, 7*, 1-14.

Colpaert, J., Cucchiarini, C., Strik, H., & Oberhofer, M. (2011, May). *Speech-enabled CALL and pronunciation training: Learning goals versus feasibility?* Paper presented at CALICO Conference, University of Victoria.

Cohen, A. (2005). Strategies for learning and performing L2 speech acts. *Intercultural Pragmatics, 2*, 275-301.

Crystal, D. (1997). *English as a global language*. Cambridge: Cambridge University Press.

Dam, L. (2000). Evaluating autonomous learning. In B. Sinclair, I. McGrath, & T. Lamp (Eds.), *Learner autonomy, teacher autonomy: Future directions* (pp.48–59). Harlow, UK: Longman.

DeKeyser, R. (Ed.). (2007). *Practicing in a second language: Perspectives from applied linguistics and cognitive psychology*. Cambridge: Cambridge University Press.

DeKeyser, R. (2010). Practice for second language learning: Don't throw out the baby with the bathwater. *International Journal of English Studies*, *10*(1), 155-165.

Duff, P. A. (2007). Second language socialization as sociocultural theory: Insights and issues. *Language Teaching*, *40*(4), 309-319.

Eisenmann, M. & Summer, T. (2012). Oral exams: Preparing and testing students. In M. Eisenmann & T. Summer (Eds.), *Basic issues in EFL teaching and learning* (pp. 415–428). Heidelberg: Universitätsverlag Winter.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics 30*(4), 474-509.

Ellis, N. C. & Bogart, P. (2007, October). *Speech and language technology in education: The perspective from SLA research and practice*. In Proceedings of the SlaTE Workshop on Speech and Language Technology in Education ISCA Tutorial and Research Workshop, Farmington, Pennsylvania.

Eskenazi, M. (2009). An overview of spoken language technology for education. *Speech Communication*, *51*(10), 832-844.

Evanini, K., Tsuprun, E., Timpe-Laughlin, V., Ramanarayanan, V., Lange, P., & Suendermann-Oeft, D. (2017, July). *Evaluating the impact of local context on CALL applications using spoken dialog systems.* In Proceedings of Computer Assisted Language Learning (CALL) Conference, Berkeley, CA. Retrieved from http://call2017.language.berkeley.edu/wp-content/uploads/2017/07/CALL2017_proceedings.pdf

Figura, K. & Jarvis, H. (2007). Computer-based materials: A study of learner autonomy and strategies. *System*, *35*, 448—468.

Gass, S., & Mackey, A. (2015). (2nd ed.) *Input, interaction, and output in SLA*. In B. VanPatten and J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 180-206). New York: Routledge.

Gijsen, L., & Colpaert, J. (2017, July). *Design of telecollaborative tasks: Where is the context?* In Proceedings of Computer Assisted Language Learning (CALL) Conference, Berkeley, CA. Retrieved from http://call2017.language.berkeley.edu/wp-content/uploads/2017/07/CALL2017_proceedings.pdf

Gilmore, A. (2007). Authentic materials and authenticity in foreign language learning. *Language Teaching*, *40*(2), 97-118.

Godwin-Jones, R. (2009). Emerging technologies speech tools and technologies. *Language Learning & Technology*, *13*(3), 4–11. Retrieved from http://llt.msu.edu/vol13num3/emerging.pdf

Golonka, E. M., Bowles, A. R., Frank, V. M., Richardson, D. L., & Freynik, S. (2014). Technologies for foreign language learning: A review of technology types and their effectiveness. *Computer Assisted Language Learning*, *27*(1), 70-105.

Grobler, C. & Smits, T. F. H. (2016). Designing a digital pedagogical pattern for improving foreign language learners' oral proficiency. *Literator, 37*(2), a1273. doi:http://dx.doi.org/10.4102/lit.v37i2.1273

Grobler, C. & Smits, T. F. H. (2017, July). *Road map for the context-sensitive redesign of a technology-enhanced speaking practice environment*. In Proceedings of Computer Assisted Language Learning (CALL) Conference, Berkeley, CA. Retrieved from http://call2017.language.berkeley.edu/wp-content/uploads/2017/07/CALL2017_proceedings.pdf

Healey, D. (2007). Theory and research: Autonomy and language learning. In J. Egbert,

& E. Hanson-Smith (Eds.), *CALL environments: Research, practice, and critical issues* (pp.377–388). Alexandria, VA: TESOL.

House, J. (2009). Introduction: The pragmatics of English as a lingua franca. *Intercultural Pragmatics, 6*, 141-145.

Johnson, W. L. (2010). Serious use of a serious game for language learning. *International Journal of Artificial Intelligence in Education, 20*, 175-195.

Kessler, G., & Bikowski, D. (2010). Developing collaborative autonomous learning abilities in computer mediated language learning: Attention to meaning among students in wiki space. *Computer Assisted Language Learning*, *23*(1), 41-58.

Kim, S. H. (2014). Developing autonomous learning for oral proficiency using digital storytelling. *Language Learning & Technology 18*(2), 20–35. Retrieved from http://llt.msu.edu/issues/june2014/action1.pdf

Kondo, S. (2008). Effects on pragmatic development through awareness-raising instruction: Refusals by Japanese EFL learners. In E. Alcón and A. Martínez Flor (Eds.), Investigating pragmatics in foreign language learning, teaching and testing (pp. 153–177). Clevedon: Multilingual Matters.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into practice*, *41*(4), 212-218. Retrieved from https://www.depauw.edu/files/resources/krathwohl.pdf

Lantolf, J. P., Thorne, S. L., & Poehner, M. E. (2015). (2nd Ed.). Sociocultural theory and second language development. In B. VanPatten and J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 207-226). New York: Routledge.

Laurillard, D. (2012). *Teaching as a design science: building pedagogical patterns for learning and technology*. New York, NY: Routledge.

Lightbown, P. (2000). Classroom SLA research and second language teaching. *Applied Linguistics*, *21*(4), 431-462.

Little, D. (1991). *Learner autonomy 1: Definitions, issues and problems*. Dublin: Authentik.

Long, M. (1991). Focus on form: A design feature in language teaching methodology. In K. de Bot, R. Ginsberg, & C. Kramsch (Eds.), *Foreign language research in crosscultural perspectives* (pp. 39-52). Amsterdam: John Benjamins.

Lyster, R., & Sato, M. (2013). Skill acquisition theory and the role of practice in L2 development. In M.G. Mayo, J. Gutierrez-Mangado, & M.M. Adrián (Eds.), *Contemporary approaches to second language acquisition* (pp. 71–92). Amsterdam: John Benjamins.

Martínez-Flor, A., & Usó-Juan, E. (2006). A comprehensive pedagogical framework to develop pragmatics in the foreign language classroom: The 6Rs approach. *Applied Language Learning*, *16*(2), 39-64.

McGraw, I., & Seneff, S. (2007). Immersive second language acquisition in narrow domains: A prototype ISLAND dialog system. *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education* (pp. 84-87). Farmington, PA: Carnegie Mellon University and ISCA Archive.

Morrow, K. (1977). Authentic texts and ESP. In S. Holden (Ed.), *English for Specific Purposes* (pp. 13-17). London: Modern English Publications.

Morton, H., & Jack, M. A. (2005). Scenario-based spoken interaction with virtual agents. *Computer Assisted Language Learning*, *18*(3), 171-191.

Morton, H., & Jack, M. (2010). Speech interactive computer-assisted language learning: A cross-cultural evaluation. *Computer Assisted Language Learning*, *23*(4), 295-319.

Morton, H., Gunson, N., & Jack, M. A. (2012). Interactive language learning through speech enabled virtual scenarios. *Advances in Human-Computer Interaction*, *1-14*.

O'Dowd, R. & Ware, P. (2009). Critical issues in telecollaborative task design. *Computer Assisted Language Learning*, *22*, 173-188.

Pino-James, N. (2013). Exploring the benefits of ACMC for speaking development. *The IALLT Journal, 43*(1), 25–63.

Ramanarayanan, V., Suendermann-Oeft, D., Lange, P., Mundkowsky, R., Ivanov, A., Yu, Z., Qian, Y., & Evanini, K. (2017). Assembling the jigsaw: How multiple open standards are synergistically combined in the HALEF multimodal dialog system. In D. A. Dahl (Ed.), *Multimodal interaction with W3C standards: Towards natural user interfaces to everything* (pp. 295–310). Springer.

Reinders, H. & White, C. (2010). The theory and practice of technology in materials development and task design. In Harwoord, N. (Ed.), *Materials in ELT: Theory and practice* (pp. 58-80). Cambridge: Cambridge University Press.

Roever, C., Fraser, C., & Elder, C. (2014). *Testing ESL sociopragmatics: Development and validation of a web-based test battery*. New York, NY: Peter Lang.

Shea, P. & Bidjerano, T. (2010). Learning presence: Towards a theory of self-efficacy, self-regulation, and the development of a communities of inquiry in online and blended learning environments. *Computers & Education, 55*(4), 1721—1731.

Seneff, S., Wang, C., & Zhang, J. (2004). Spoken conversational interaction for language learning. *Proceedings of InSTIL/ICALL Symposium* (Paper 036). Venice, Italy: ISCA Archive.

Steyn, J. & Grobler, C. (2016). *Papotons!* [computer software], NWU, Potchefstroom.

Strik, H., Cornillie, F., Colpaert, J., van Doremalen, J., Cucchiarini, C. (2009).

Developing a CALL system for practicing oral proficiency: How to design for speech technology, pedagogy and learners. In *Proceedings of SLaTE-2009,* Wroxall Abbey.

Stracke, E. (2007). Spotlight on blended language learning: A frontier beyond learner autonomy and computer assisted language learning. In *Proceedings of the Independent Learning Association 2007 Japan conference: Exploring theory, enhancing practice: Autonomy across the disciplines*. Chiba, Japan: Kanda University of International Studies.

Sun, Y-C., & Yang, F-Y. (2015). I help, therefore, I learn: service learning on Web 2.0 in an EFL speaking class. *Computer Assisted Language Learning, 28,* 202-219. http://dx.doi.org/10.1080/09588221.2013.818555

Swain, M., & Watanabe, Y. (2013). Languaging: Collaborative dialog as a source of second language learning. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1-8). Oxford: Wiley-Blackwell.

Sydorenko, T. (2011). Exploring the potential of rehearsal via automatized structured tasks versus face-to-face pair work to facilitate pragmatic and oral development. Michigan State University Doctoral dissertation. Dissertation Abstracts International 3464977.

Sydorenko, T. (2015). The use of computer-delivered structured tasks in pragmatic instruction: An exploratory study. *Intercultural Pragmatics, 12*(3), 333-362.

Sydorenko, T. (2016). *Development of adaptive computer-simulated conversations with video support: Lessons learned*. Paper presented at CALICO, East Lansing, MI.

Sydorenko, T., & Tuason, G. H. (2016). Noticing of pragmatic features during spoken interaction. In K. Bardovi-Harlig and C. Félix-Brasdefer (Eds.), *Pragmatics & Language Learning, Vol. 14* (pp. 233-264). Hololulu, HI: University of Hawai'i, National Foreign Language Recource Center.

Sydorenko, T., Daurio, P., & Thorne, S. (2018). Refining pragmatically-appropriate oral communication via computer-simulated conversations. *Computer Assisted Language Learning*, *31*, 157-180.

Taguchi, N. (2012). *Context, individual differences, and pragmatic competence*. Toronto: Multilingual Matters.

Taguchi, N., & Sykes, J. (Eds.) (2013). *Technology in interlanguage pragmatics research and teaching*. Amsterdam: John Benjamins.

Timpe-Laughlin, V., Evanini, K., Green, A., Blood, I., Dombi, J., & Ramanarayanan, V. (2017). Designing interactive, automated dialogs for L2 pragmatics learning. In V. Petukhova and Y. Tian (Eds.), *Proceedings of the 21st Workshop on the Semantics and Pragmatics of Dialog* (pp. 143-152).

Timpe-Laughlin, V., Wain, J., & Schmidgall, J. (2015). Defining and operationalizing the construct of pragmatic competence: Review and recommendations. *ETS Research Reports*. Princeton, NJ: Educational Testing Service.

van Compernolle, R. A. (2011). Developing second language sociopragmatic knowledge through concept-based instruction: A microgenetic case study. *Journal of Pragmatics, 43,* 3267-3283.

van Doremalen, J., Boves, L., Colpaert, J., Cucchiarini, C., & Strik, H. (2016). Evaluating automatic speech recognition-based language learning systems: A case study. *Computer Assisted Language Learning, 29*, 833-851.

Wik, P., & Hjalmarsson, A. (2009). Embodied conversational agents in computer assisted language learning. *Speech communication*, *51*(10), 1024-1037.

Winne, P.H. (2005). A perspective on state-of-the-art research on self-regulated learning. *Instructional Science, 33*, 559—565.

Yen, Y-C., Hou, H-T., & Chang, K. E. (2015). Applying role-playing strategy to enhance learners' writing and speaking skills in EFL courses using Facebook and Skype as learning tools: A case study in Taiwan. *Computer Assisted Language Learning*, *28*, 383-406.

Appendices

Appendix A

Additional Details on *Papotons!* Learning Cycle and Study Procedures

The *Papotons!* practice environment is based on Laurillard's (2012) Conversational Framework (see Grobler & Smits, 2016) and accordingly consists of several stages within a learning cycle. A cycle begins with in-class preparatory activities on, for instance, greeting someone, introducing oneself and spelling one's name. The program (see table below) runs concurrent with classes, which are concluded by a one-on-one question-and-answer style oral exam with the teacher, focusing on the outcomes stated at the beginning of the cycle.

Table A1. *Papotons!* and LMS learning cycle.

1. LMS: listening to model dialogue and doing short quiz	A recording of a conversation models the language to master by the end of the learning cycle. Simple multiple choice questions in the foreign language about the content of the recording ensure that students understood the recording.
2. *Papotons!*: taking part in a simulated conversation with self-assessment	*Papotons!* allows to create a dialogue with interview-type questions to be listened to and students' answers to be recorded (each answer can be re-recorded as many times as needed before being saved). Students then do a self-assessment of their performance and answer a reflection question on an aspect of the program, after which the work is submitted for feedback and assessment.

3. LMS: receiving audio feedback (+ assessment)	Teacher gives audio feedback on each individual answer recorded (using predetermined assessment criteria). Students listen to the integrated recording (i.e., questions, answers and feedback), after which they complete a short quiz (on the LMS) on aspects of the feedback (e.g. how many times they listened to it, which aspects were most helpful).
4. mobile phones, LMS: producing videos in pairs and commenting on peers' videos	Students make a short video in pairs to illustrate their mastery of the outcomes for the learning cycle: they create a scenario where they have a conversation providing personal information and asking questions to learn about the other person (name, date of birth, hobbies, etc. depending on the outcomes). Peers comment on one of the videos using predetermined criteria (e.g. fluency, accent, pronunciation, intelligibility, etc).

Note. LMS = Learning Management System.

The *Papotons!* practice environment offered a solution to the problem of limited opportunity for students to express themselves in and outside class, and to get meaningful feedback. The technological development of the program, however, was constrained by financial limits (absence of funding for elaborate visual features or development as a mobile phone app), by data restrictions (need for downloadable software, problem of unsteady or lacking home internet connection) and by the targeted language level (avatars would lack authentic lip synchronization, which beginners are prone to focus on).

In the 2017 academic year, the total number of enrolled French students was 67. 53 students participated in the first practice cycle (54 took the subsequent oral exam) and 46 students in the second (47 took the second oral exam), meaning that between 68 and 80% of students took part in both the in- and out of class learning activities. 29 videos were submitted during practice cycle 1 and 24 in the course of the second (i.e., not all students worked in pairs). The *Papotons!* quiz with which students evaluated the program after receiving teacher feedback (step 3 in Table A1 above) was done by 31 students (58.4% response) at the end of the first practice cycle and by 24 (52.1%) after the second. For the (also voluntary) SDLI survey, only the data from the 21 (identical) students that filled out a questionnaire for each of the practice cycles are used in this article.

Appendix B

Additional Details on *SimCon* Data Analysis

When analyzing the data for learning outcomes, a microgenetic approach, which is used for tracking qualitative changes in cognition and performance over a short period of time (van Compernolle, 2011) was adopted. Thus, changes learners made in oral requests for a letter of recommendation within a two-and-a-half-hour time frame were examined. If learners' changes in oral production became more congruent with the system's models (i.e., student videos), we considered those to be positive learning outcomes as the model student videos had been vetted as pragmatically appropriate.

Studies examining structured practice of pragmatic routines indicate that learners can focus on both linguistic form and the content of their responses (Sydorenko, 2015; Sydorenko & Tuason, 2016). Changes made by learners were thus put into two categories: form or content. Evidence of focus on form was operationalized as learners' use of expert language from the model videos in posttest simulations 5 and 6. Following Sydorenko (2015), changes in any forms, including not only pragmatically appropriate expressions like *I really appreciate it*, but also vocabulary and morphosyntax, were examined because linguistic competence and pragmatic knowledge are connected (e.g., Roever, Fraser, & Elder, 2014). For example, if learners used the video-modelled expression *relevant information* in posttest but not in pretest simulations, it was counted as a form change. Following Sydorenko (2015) and Kondo (2008), evidence of change in content was determined by participants' use of content (typically, sociopragmatic strategies) similar to the model videos. For example, if a participant added an explanation for their late request in the posttest, and if an explanation was present in model videos, this was coded as change in content.

In terms of form, common words or expressions adopted from models were: *thank you for your consideration*, *unfortunately*, *found out*, and *relevant information*. The three instances of personal form changes were linguistic reformulations of the same request strategies. For example, participant FB9's original explanation was "the scholarship unfortunately is really important for you- me" and on the second try this explanation was changed to the linguistic form "it's very crucial for me."

With regard to content, appropriation from models was generally in the form of sociopragmatic strategies; particularly common were:

- appealers (i.e., an effort by a speaker to appeal to the hearer's benevolent understanding), such as "It's important for me in the finance way;"
- disarmers (i.e., when "the speaker tries to remove any potential objections the hearer might raise upon being confronted with the request" (Blum-Kulka, House, & Kasper, 1989, p. 287)), such as "I really understand your situation and I know it's pretty short notice;"
- grounders (i.e., reasons for the request), such as "It's very related with the class that we are taking now."

(Definitions of strategies commonly used in cross-cultural requests were drawn from Sydorenko & Tuason, 2016; Blum-Kulka et al., 1989; Taguchi, 2012).

With regard to cultural notions, some participants realized that when talking to instructors in the U.S., office hours should be utilized, particular greetings should be used, and instructors seem to be non-judgmental and very approachable. For many, the appropriate timeframe for the request, influenced by differences in L1 and L2 cultures, became particularly salient. Participants also reported observing the organization of request sequences, such as providing a balanced level of detail (not too much, not too little) before making the request.

[1] Intrinsic feedback is feedback that is meaningful, readily available and formative. It is inherent to and/or forms part of the learning activity proper and "takes the form of a natural or authentic consequence of the action in relation to the intended goal, from which the learner can work out how to improve their action without teacher intervention" (Laurillard, 2012, p. 55).

[2] However, less than 10% of the students present at contact sessions in language laboratory sought technical support, likely due to the fact that sufficient technical support was provided face-to-face in language laboratory. In 2017, no students appealed to the teacher/facilitator for content support as the teacher was not allowed to give such a support in our test environment of that year. For this kind of support, students had to recur to the model dialogue or their textbooks, and could ask their fellow students (which they did).

[3] According to *Cohen's d* guidelines, .2 = small, .5 = medium, and .8 = large.

[4] Like Morrow (1977, p. 13), we defined authenticity as follows: "An authentic text is a stretch of real language, produced by a real speaker or writer for a real audience and designed to convey a real message of some sort." However, we caution readers that many other definitions of authenticity exist (see Gilmore, 2007).

[5] http://halef.org

[6] A prototype version of this *HALEF* task, along with a few other sample *HALEF* tasks, is available at the following website: www.englishtasks.org.

[7] LOTS and HOTS are represented in Krathwohl (2002) as a continuum of skills rather than two distinct categories. However, for ease of reference, in this paper we conceptualize the first three

types of cognitive processes (remember, understand, apply) as LOTS and the last three (analyze, evaluate, create) as HOTS.