



An embodied conversational agent coach to support societal participation learning by low-literate users

Dylan G. M. Schouten¹ · Agnes A. Deneka² · Mariët Theune² · Mark A. Neerincx^{1,3} · Anita H. M. Cremers^{3,4}

Accepted: 29 November 2021 / Published online: 2 April 2022
© The Author(s) 2022

Abstract

People of low literacy could benefit from automated support when learning about societal participation. We design an Embodied Conversational Agent (ECA) ‘coach’ that can provide effective learning support to low-literate learners, develop a prototype virtual learning environment, and evaluate this prototype with low-literate end users. First, we inventory the learning support benefits of ECA coaching. Second, we update existing requirements to better specify functional demands for the coach ECA. Third, we write use cases and develop the prototype. Finally, we evaluate the prototype with low-literate users in a mixed-method within-subjects experiment. Results show that the coach influences the subjective learning experience: Participants report higher positive affect, higher user-system engagement, and increased self-efficacy regarding online banking. These results particularly apply to the domain of challenging information skills exercises. Caveats apply: One of four exercises was significantly more difficult than the other three; and coach support rules were not clearly formalized.

Keywords Embodied conversational agents · Socio-cognitive engineering · Virtual learning environment · Requirements engineering · Societal participation · Low literacy

1 Introduction

In modern information societies (such as the Netherlands), the societal participation of low-literate citizens is problematically low [20]. Limited information skills (reading and writing) and communication skills (speaking and understanding) can lead to problems with societal participation. These problems can be cognitive in nature (such as a lack of these skills, and societal knowledge and experience), but also affective (fear, shame, and low self-efficacy, an individual’s task- and context-specific judgment of their own capabilities, cf. [5, 89]) or social (low motivation and desire to learn, or low trust in teachers and other learners, cf. [36, 39]). We can address them by designing information and communication skills training that is grounded in relevant

real-life societal participation scenarios, so-called crucial practical situations [55, 56]. To this end, we are designing the virtual learning environment VESSEL: A Virtual Environment to Support the Societal participation Education of Low-literates [89]. In VESSEL, learners will perform interactive exercises situated in the domain of societal participation, while the system provides learning support by addressing the combined cognitive, affective, and social spectrum of learning problems that low-literates experience. We predict that learning in VESSEL will result in higher *learning effectiveness*, which we define as consisting of *learning accessibility* (there should be no practical or emotional barriers for the learner to start learning, cf. [94–96]), *learning experience* (the learners’ skills, needs, and wishes should be incorporated throughout the learning, cf. [24, 60, 96]), and *learning outcomes* (the learning should aim to reach meaningful and desired goals, cf. [7, 60, 94, 108]) [88]. By applying cognitive, affective, and social perspectives, we identify nine concrete system objectives. Learning accessibility can be increased by lowering (1) cognitive, (2) affective, and (3) social barriers to learning. The learning experience can be made more (4) cognitively achievable, (5) affectively positive, and (6) socially engaging. Finally, the following important learning outcomes can be reached:

✉ Dylan G. M. Schouten
dyan.schouten@gmail.com

¹ Technische Universiteit Delft, Delft, Zuid-Holland, Netherlands

² Universiteit Twente, Enschede, Overijssel, Netherlands

³ TNO Soesterberg, Utrecht, Soesterberg, Netherlands

⁴ Hogeschool Utrecht, Utrecht, Netherlands

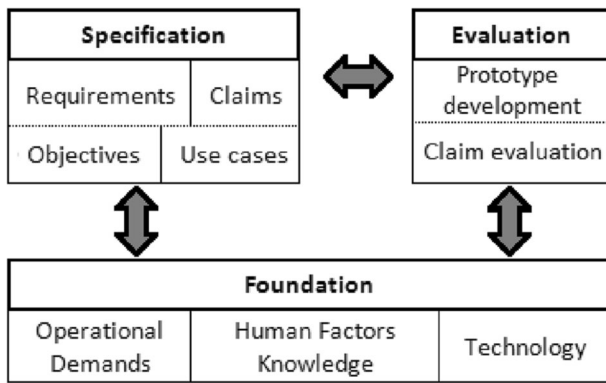


Fig. 1 Socio-Cognitive Engineering method [74, 76]

Learners can (7) train applied information and communication skills and gain practical experience, (8) raise their self-efficacy, and (9) become more motivated to participate in society independently.

We use the Socio-Cognitive Engineering method (SCE, see Fig. 1) to design VESSEL. The SCE method integrates operational demands (describing the system’s context-of-use), human factors knowledge (describing theory relevant to user-system interactions), and technology (describing current and envisioned technology drivers and constraints) into an iterative software design process [74–76]. Relevant data about these operational demands, human factors, and technology are collected in a theoretical foundation. This foundation is used to derive a system specification, consisting of system objectives (the general operational or domain goals of the envisioned system), functional requirements (the system’s intended functionality), claims (hypotheses about how the requirements help reach system objectives), and use cases (sequences of actions that describe how the system results in valuable outcomes for particular actors, cf. [1]). This specification can then be developed into a prototype, which is used to experimentally evaluate the claims.

Earlier work [88, 89] has resulted in a high-level VESSEL specification, consisting of a requirements baseline with eight requirements (see Table 1) and claims that connect these requirements to the nine system objectives of learning effectiveness. This requirements baseline is theoretically supported [89]: The requirements in the specification were derived from theories of adult learning (andragogy [51–53], transformative learning [65, 66], constructivism [10, 45], and e-learning [31, 46, 47]) and theories on computer-supported learning that highlight the value of information provision, worldwide communication, interactivity, and use of gaming principles/gamification [85]. The baseline is also empirically grounded [88]: The requirements were refined by applying grounded theory [35] to qualitative data obtained through workshops, focus groups, and cultural probes (a qualitative data collection method wherein

participants use provided recording tools, such as cameras, notepads, and sound recorders, to provide insight into their daily lives, cf. [32, 33]) used with low-literate participants. However, the requirements and claims in this specification have not yet been practically evaluated. As such, the next steps in our design and development process should be the ‘prototype development’ and ‘claim evaluation’ steps in the evaluation phase (Fig. 1). During prototype development, we must translate the generic requirements baseline into use cases, low-level claims, and functional VESSEL prototypes. We then use these prototypes during claim evaluation to experimentally test the validity of the claims with low-literate end users. The experimental outcomes of this evaluation phase can then be used to update the foundation and refine the specification, iteratively improving the VESSEL design.

The high-level specification affords a range of possible technological implementations, each meeting certain requirements in certain ways. We choose to design VESSEL as an autonomous rules-driven Embodied Conversational Agent (ECA) coach that helps low-literate learners with situated interactive exercises by offering cognitive, affective, and social learning support. Figure 2 presents the envisioned VESSEL system setup, showing the ECA coach and exercise elements. Here, exercises are scenario-based training (cf. [82]) situated in crucial practical situations important to low-literate learners. ECAs are “*anthropomorphic interface agents*” [12, p. 1] that can directly interact with system users. We expect that an ECA coach implementation of VESSEL has theoretical and empirical benefits. Example benefits include the following: ECAs can adapt their looks and behaviours easily, allowing

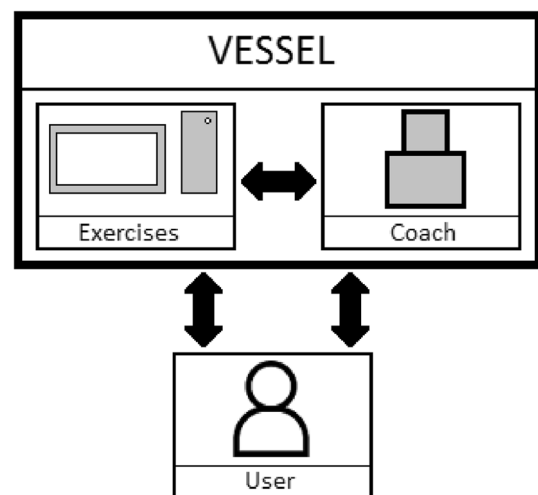


Fig. 2 VESSEL system setup. Three double-sided arrows indicate user-system interactions. Bottom left arrow: The user performs exercises. Bottom right arrow: The coach monitors the user’s actions, and interacts with the user by giving feedback and support. Top arrow: The coach monitors exercise state and changes support as appropriate

them to match the demands of different training scenarios or the needs and wishes of different users; ECAs inherently afford natural language and spoken dialogue, making them easier to communicate with especially for users who struggle with text and reading; ECAs can add a social presence to exercises; particularly in a coaching role, they can serve as a focal point for user support, allowing users to naturally ask questions and request help. Finally, an ECA coach matches the support desires of low literates: In [88], we show that low-literate learners strongly prefer personalized ‘human’ support over ‘computer’ support. An anthropomorphic ECA puts a human face on the computer system, allowing low-literate learners to access the benefits of automated support, thereby enhancing learning effectiveness. However, to the best of our knowledge, little experimental validation of the effectiveness of ECA coaching with low-literate societal participation learners currently exists.

Following the SCE methodology, we aim to address this larger problem by carrying out multiple design and evaluation cycles. In this paper, we aim to design and develop a VESSEL prototype consisting of situated interactive exercises and an ECA coach that provides cognitive, affective, and social learning support and evaluates this prototype with low-literate learners. The prototype developed in this work will be a proof-of-concept, used in a Wizard-of-Oz experiment (i.e. controlled by a human ‘wizard’ instead of a computer, cf. [64]) to investigate both how we can best translate the existing VESSEL specification into an ECA coach-supported virtual learning environment, and whether or not the idea of ECA coach support for low-literate learners provides the envisioned learning benefits as described above: Better learning accessibility, an improved learning experience, and more success at reaching important and meaningful learning outcomes. The question of whether or not low-literate learners will be able to benefit from this prototype is non-trivial, as low-literate learners are known to struggle with accessing and using complex technology due to cognitive, affective, and social barriers [88, 90]. Additionally, Kramer et al. [54] highlight that many ECA studies underreport the actual ECA design process and argue for studies that “*open the black box*” (p.8) and clearly articulate the methods, objectives, and assumptions that go into this design. Ter Stal et al. [101] similarly report a dearth of clear guidelines for and taxonomies of ECA design features. Consequently, the comprehensive design, development, and evaluation of a proof-of-concept ECA coach that provides cognitive, affective, and social learning support meaningfully integrated into exercises are a unique and interesting contribution. A complementary study by Schouten, Venneker et al. [91] has zoomed in on the affective and social support contributions of a

different prototype, further exploring the boundaries of this problem space.

The above yields two research questions:

- *Q1. Design* How can we create an ECA coach that provides effective cognitive, affective, and social learning support to low-literate learners doing situated interactive exercises in a virtual learning environment?
 - *Q1a.* In what ways can an ECA coach provide cognitive, affective, and social learning support?
 - *Q1b.* Which functionalities, interaction methods, and appearances should an ECA coach have to effectively provide this learning support in a virtual learning environment?
- *Q2. Evaluation* Does this support-providing ECA coach increase learning effectiveness for low-literate learners working with VESSEL, compared to low-literate learners working with VESSEL but not receiving coach support?

The first research question is answered in four steps. First, we update our SCE foundation (Fig. 1). We update technology by explaining the potential benefits that ECAs in general, and an ECA coach specifically, can offer to our VESSEL design. We update human factors knowledge by incorporating theory that describes how an ECA coach could offer cognitive, affective, and social learning support. We update operational demands by designing the situated interactive exercises that make up the educational content of VESSEL. Second, this updated foundation is used to refine the requirements in the specification. Third, we translate this refined specification into practical use cases, to make explicit how the prototype should work, what effects we expect our ECA coach to have, and how these effects can be measured. Fourth, we create the prototype and describe it in terms of functionality, interaction methods, and appearance. To answer the second question, we experimentally evaluate the learning effectiveness impact of the coach with low-literate learners. We claim that a support-providing ECA coach will raise VESSEL’s learning effectiveness for low-literate learners by improving the system’s cognitive, affective, and social learning accessibility, learning experience, and learning outcomes.

The structure of this paper is as follows. Section 2 provides additional background information on the demographic of low-literate learners, and on the current state of learning support aimed at this demographic. Section 3 shows the updated VESSEL foundation. In section 4, the specification requirements are refined, and use cases are derived and written. Sections 5, 6, and 7 present the evaluation process. In Sect. 5, the prototype is described in terms

of functionality, interaction methods, and appearance. Section 6 describes the experiment created to evaluate the effectiveness of the learning support provided by the prototype's ECA coach. Section 7 presents the results of the evaluation. Finally, Sect. 8 presents conclusions, discussion of findings, and directions for future work.

2 Background

2.1 People of low literacy

People of low literacy (or *low-literate people*) are defined as adult people whose mastery of reading, writing, speaking, and understanding/comprehension skills are limited to a degree that they cannot *independently participate* in society. The OECD (Organisation for Economic Cooperation and Development) defines literacy as the ability to use printed and spoken information in the pursuit of one's daily and overall goals [78]. Low literacy can specifically be measured against *crucial practical situations*: The set of behaviours that a person must be able to carry out in order to be able to participate independently [55, 56]. Examples of these crucial practical situations include banking, renting a living space, grocery shopping, and communicating with neighbours [55, 88].

As the elements of low literacy are highly culturally dependent (owing to, e.g. different norms, expectations, governmental institutions, and crucial practical situations, cf. [88]), we focus on low literacy in the Netherlands, the country where this study is situated. Following the most recent definitions provided by the Dutch Court of Audit (*Algemene Rekenkamer*), around 2.5 million Dutch people are considered low-literate [2], including 1.8 million people in the labour force (i.e. aged between 16 and 65). This represents an increase over the 2012 *Programme for the International Assessment of Adult Competencies* (PIAAC) survey, which reported 1.3 million low-literate Dutch people [37, 40]. Gubbels et al. [41] report that almost a quarter of Dutch 15-year olds struggle with language mastery and literacy issue, further cementing the widespread nature and persistence of this issue. Low literacy is similarly heterogeneous among other demographic angles: Low literacy can be found among both native Dutch speakers and non-native migrants, among both men and women, and among a variety of educational backgrounds [87]. A more comprehensive look into the demographic makeup of Dutch low literates can be found in Schouten, Smets et al [89] and De Greef, Segers & Nijhuis [38].

2.2 Existing learning support

Learning support for Dutch low literates is provided primarily in the form of adult language and integration classes, which are made available to interested learners at regional education centres, libraries, volunteer centres, and private institutions [11, 100]. In these classes, which focus equally on language acquisition, practical skills training, and knowledge of society, students discuss and practice crucial practical situation exercises with the support of teachers and peers [34, 55]. This classroom-based approach to training is kept as the gold standard because it meaningfully applies scenario-based learning [80, 82] to the actual practice of societal participation that low-literate people struggle with, and because low-literate learners put enormous value into support from both trusted authority figures (i.e. teachers) and people that they perceive to be in the same situation (i.e. student peers) [88].

However, Schouten [87] highlights that classroom learning can meaningfully be supplemented with computer-based learning, citing three broad reasons: First, computer-based learning can improve learning accessibility for learners who cannot access the classroom easily. Second, computer-based learning has great inherent potential for individualization (as learning software is easily adaptable, cf. [6, 80]); individualized learning and support are very valuable for increasing learning effectiveness [69, 97], especially for a heterogeneous learner group like low-literate learners [89]. Finally, computer-based learning enables the use of learning scenarios that would be impractical or impossible to train in a classroom setting (e.g. scenarios that have an element of real risk, or that require expensive tools). Schouten, Smets et al. [90] assess the current practice of learning support software aimed at low-literate learners and use this to establish a set of eight design requirements for learning support software that emphasizes these benefits. They describe that a virtual learning environment (VLE) could potentially be an effective way of meeting these requirements. These design requirements (further refined in [88] and shown in Table 1) form the basis of VESSEL: A VLE aimed at effectively supporting low-literate learners who want to improve their societal participation, designed to make the benefits of computer-based learning accessible to low-literate learners and to be used in concert with existing learning solutions.

3 Foundation

3.1 Technology: embodied conversational agents

Embodied Conversational Agents (ECAs) are a subclass of Intelligent Virtual Agents: Autonomous software programmes that can interact with humans and other agents

[12]. ECAs extend from traditional intelligent agents by being ‘embodied’ as animated virtual characters in a virtual environment. Being embodied has consequences for agent appearance and behaviour. In contrast to non-embodied agents, ECAs can be judged on their appearance; particularly when ECAs look human-like, humans evaluate it on appearance factors such as sex, age, ethnicity, and dress style. Studies suggest that humans judge ECA characters on the same qualities as they do other humans, such as similarity to themselves [3, 70, 105], attractiveness [49, 72], and cultural appearance stereotypes [3, 4, 109]. In addition, ECAs can use not only verbal communication behaviours (e.g. speech and natural language understanding), but also nonverbal behaviours (e.g. body language, gesturing, facial expressions, and gaze direction [23, 54, 101]). ECAs can be designed to behave as social actors: Potential possibilities include recognizing and responding to verbal and nonverbal input from humans, taking part in ongoing discussions, paying attention, and using conversational functions like turn-taking [12, 23, 28, 62]. This lets humans react to the social cues and behaviours of ECAs as if they were human conversation partners [50, 73, 84, 98].

Because ECA behaviours and appearances can be adjusted [6, 80], ECAs can be used to fulfil a variety of roles in a virtual environment. For instance, Bickmore et al. [14] adapted the ethnicity of a virtual nurse character to better match different user demographics, increasing user-system satisfaction. Prior studies have shown the potential effectiveness of using ECAs in the role of a digital coach. Lane et al. [57] report on an ECA coach that increased users’ willingness to attempt challenging programming problems and their self-efficacy in computer science education. Coaching ECAs developed by both Bickmore et al. [16] and de Rosis et al. [86] were effective in changing user behaviour patterns. Shamekhi et al. [92] show that an ECA coach focussed on teaching self-care was appreciated and accepted by spinal cord injury patients, with participants suggesting that this approach could be valuable particularly for adults dealing with ‘sensitive topics’. Hudlicka [44] shows that even when users express negative opinions on an ECA coach’s affective and social realism (i.e. the ECA’s ability to conduct natural conversations and come across as a ‘real’ person), the coach’s interactive feedback and support are still valued, and the coach still supports users in implementing a meaningful meditation practice routine. Finally, Kramer et al. [54] provide a scoping literature review of the use of coaching ECAs in physical health domains. They report that while no significant increase in user health literacy is found, ECA coaches do increase user motivation to apply health measures, user identification of preconception risks, and system usability.

Because of the social interaction options, individualization potential, and learning support outcomes described above, an ECA fulfilling the role of a digital coach in

VESSEL could be effective in supporting the cognitive, affective, and social issues of low-literate learners. Cognitively, digital coaches in general can help learners reach stated learning goals: Bickmore et al.’s [16] health counsellor increased physical activity and fruit and vegetable consumption, and Veletsianos and Miller [104] show that learners deeply engage and converse with a digital coach, increasing learning. ECA digital coaches in particular can provide individualized learning support by using scaffolding to structure their verbal/textual feedback and by using multimodal media for support [63]. De Rosis et al. [86] show that users converse with and learn from an ECA for health promotion with adaptive dialogue, and Miao et al. [67] show that a scaffolding-based ECA coach is both technically feasible and accepted by learners. Affectively, ECA digital coaches improve the affective experience of situated, interactive learning exercises: Shaw et al. [93] describe how the embodied, human-like nature of an ECA can emotionally benefit learners in situated exercises, while Lester et al. describe that “*the very presence of an animated agent in an interactive learning environment - even one that is not expressive - can have a strong positive effect on student’s perception of their learning experience*” [59, p. 6]; they call this the *persona effect*. Also, socially, digital coaches can be seen as ‘friends’ and trusted mentors in a learning system [13, 86], forming a long-term relationship of trust between learner and coach [15, 71]. Ter Stal et al. [101] report that social relational agents are seen as more likeable, caring, and trustworthy, particularly if the ECA shares information about itself with the user (cf. [48]). ECA digital coaches can use nonverbal behaviours and appearance factors, such as similarity attraction, to form these relationships quickly and strongly [70]. Digital coaches also enhance engagement and learning in a virtual learning space, by acting as conversation partners that human users will genuinely talk to [104].

3.2 Human factors knowledge: learning support

To support learners with cognitive, affective, and social issues, the ECA coach must be able to offer cognitive, affective and social support. Cognitive support is operationalized in VESSEL as *scaffolding*. Scaffolding is a learning support technique that focuses on providing the right amount of support to learners at the right time. Support is first increased to the level that the learners need to progress and then gradually decreased over time [83]. This way, “*students are encouraged to develop their own creativity, motivation, and resourcefulness*” [103, p. 652]. The coach can use verbal scaffolding techniques [27, 29] by offering exercise-specific explanations and hints; this helps learners understand the learning content and successfully complete the exercise. Affective support is operationalized as *motivational interviewing*. Motivational interviewing is a

counselling technique aimed at leveraging intrinsic motivation to enact behavioural change [68]. The motivational interviewing techniques help learners to feel good about the learning process, and to reframe and solidify positive self-efficacy information (cf. [17, 30, 61]). The coach can use motivational interviewing techniques by offering exercise-specific feedback, eliciting self-reflection, and applying social persuasion to raise learner self-efficacy [99]. Social support is operationalized as *small talk*, which is a cornerstone of building trust. Trust is an important element of the learning process [12, 21], as it makes learners more receptive to the coach's suggestions and motivates learner persistence. Small talk leads to the building of trust by increasing coordination between speaking partners, establishing common ground, and helping to keep the conversation at a safe level of depth, thereby avoiding 'face threat' [22]. The coach can apply these categories in exercise-specific small talk.

3.3 Operational demands: exercises

For this prototype, we require a set of situated exercises that covers a range of possible cognitive, affective, and social issues that low-literates can encounter in daily life. We draw two exercise scenarios from the list of crucial practical situations: 'using online banking' (online banking), and 'requesting a new passport at a city hall service desk' (service desk). These two exercise domains test different skill sets: online banking tests reading and writing skills, while service desk tests speaking and comprehension. Furthermore, we apply two difficulty levels to each scenario, 'Easy' and 'Hard', resulting in four exercises: *Easy online banking*, *Hard online banking*, *Easy service desk*, and *Hard service desk*. This step has two purposes: Firstly, using four different exercises will provide a larger and more varied range of data than using two, while keeping pairs of exercises in the same domain (i.e. two online banking and two service desk exercises) enables more meaningful direct comparison of the outcomes. Secondly, this setup more accurately mirrors the participation experiences of people of low literacy, who can encounter both simple and difficult challenges in any given domain [88]. This allows us to compare the difference in practical experience between 'easy' and 'hard' situations and evaluate the types and amounts of support that are needed for each. All four exercises are designed to incorporate the ECA coach.

We determine which cognitive, affective, and social challenges are likely to appear in each of these four exercises, and what level of information and communication skills will be needed, using the Societal Participation Experience of Low-Literates (SPELL) model from Schouten, Paulissen et al. [88] and domain-specific literature. In the online banking exercises, the user must transfer money from a personal account to a web shop. These exercises are designed using

Bayles' [8] overview of critical online banking usability factors, and Nielsen's [77] and Leavitt & Schneiderman's [58] general usability guidelines. The difficulty between *Easy online banking* and *Hard online banking* is changed by raising/lowering the usability and user-friendliness of the websites: The *Easy online banking* website is less complex, less information-rich, and easier to navigate than the *Hard online banking* website. Visual appearances were based on examples of real-life online banking websites (see Fig. 3).

In the service desk exercises, the user must speak to a city hall employee to report the loss of a passport. This city hall employee is presented as an ECA character. The difficulty between *Easy service desk* and *Hard service desk* is changed in two steps by presenting the *Easy service desk* ECA as more friendly and polite than the *Hard service desk* one. First, we use De Jong et al.'s [25] overview of social demeanour and politeness effects to write dialogue for the ECAs. De Jong et al. provide politeness ratings for 21 dialogue tactics, ranging from imperative requests ("Do this for me") to apologetic speech ("I'm sorry, could you please do this for me"); using this overview, we write friendly and polite dialogue for the *Easy service desk* ECA, and curt and impolite dialogue for the *Hard service desk* ECA. Second, we give the ECAs different appearances: This will help learners distinguish between the two characters, reinforcing the idea that one character is a polite person, while the other one is mean. The *Easy service desk* ECA is given a friendly appearance, while the *Hard service desk* ECA is given an unfriendly appearance.

To increase the likelihood that players interpret the visual appearances of the conversation partner ECAs as 'friendly' and 'unfriendly', these appearances are taken from a pre-study [26], in which eight low-literate participants (in groups of two) were asked to rate a set of twelve ECA characters of diverse age, ethnicity, gender presentation, and dress style; the literature currently does not show clear consensus on which ECA designs are preferred in which situations [101, 102], necessitating this approach. We expected that participants would prefer those ECAs that were similar to them, and dislike ECAs that were dissimilar, based on Moreno & Flowerday [70]. ECAs were drawn from Brinkman et al.'s [19] 'Virtual Reality Exposure Therapy (VRET)' virtual environment. Participants were given paper pictures of a service desk setting (Fig. 4) of the twelve ECAs, and they were asked to 'select the four characters you would like to have as a conversation partner in this setting and order these four from best to worst'. They were then asked to 'select and order the four characters you would least like to see', and finally, to fill out the ordering with the last four characters; the process was done in three steps to avoid overloading participants. The eight obtained orderings were evaluated to see which characters were considered 'best' or 'worst' most often. All participants strongly disliked one particular

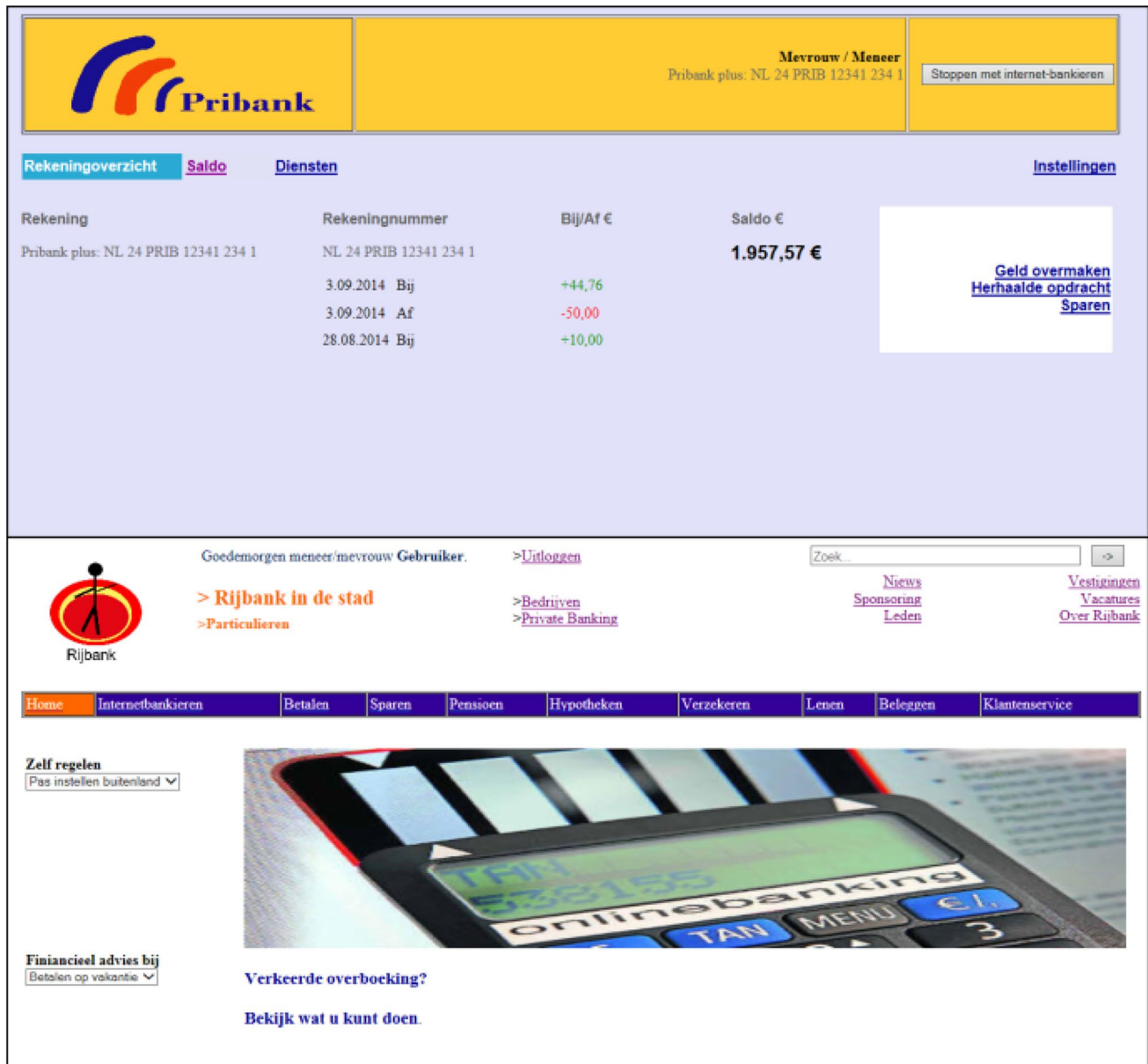


Fig. 3 Online banking exercise websites. Left: Easy online banking. Right: Hard online banking

ECA; this ECA was chosen for the *Hard service desk* exercise. Three similar-looking ECAs shared the ‘best’ spot; we selected one of these for the *Easy service desk* exercise.

Both ECAs share a small number of visual commonalities. They have one set facial expression. They open and close their mouth on a set pattern while speaking, regardless of speech content. They go through one simple ‘idle’ animation loop, swaying left and right slightly while sitting on a chair; apart from this, they employ no other body movements or gestures of any kind. Both ECAs can be seen in the context of the service desk exercises in Fig. 4; in both cases, the service desk background is the same static image.

For each exercise, written instructions are provided on-screen. In the online banking exercises, the instructions show the task (to transfer money to another account) and the information necessary to complete it: Recipient name and bank account number, and money amount. In the service desk exercises, the instructions only show the task. All exercises have been designed with a 6-minute time limit, in order to define ‘success’ (the exercise is correctly completed within 6 min) and ‘fail’ (the exercise is completed incorrectly or not within 6 min) states for the exercise. When the limit is reached, the exercise must be stopped. This 6-minute limit is based on cognitive walkthrough of the exercises and practical considerations: A shorter limit would not give participants enough



Fig. 4 ‘Conversation partner’ ECAs, shown inside the virtual environment used for the service desk exercise. Left image: Easy service desk exercise ECA. Right image: Hard service desk exercise ECA

time to reasonably do the exercises, while a longer limit would inflate the time footprint of the envisioned experimental study (see Sect. 6).

4 Specification

Following Fig. 2, we *refine* the generic ‘VESSEL’ requirements. This means that for each existing requirement, we create new, more detailed requirements that zoom in on one or both of the system’s two core components: The *exercises*, and the *coach*. When working with a large requirements baseline, careful choices must be made about which requirements to test in which configuration [107]. Since we are building a single-user prototype, we choose for the time being to discard the requirement *R4. Collaboration*, as it demands a prototype that supports multiple users at once. The remaining seven generic requirements are refined, resulting in a set of coach-specific and exercise-specific requirements. Table 1 shows the new requirements baseline. An in-depth description of the refinement process can be found in Appendix A. In addition, two use cases have been created to demonstrate the envisioned optimal way that a user would interact with VESSEL. These use cases can be found in Appendix B.

5 Evaluation: prototype development

Functionality. The prototype consists of the ECA coach that offers cognitive, affective, and social learning support as described in Sects. 3.1 and 3.2, and the exercises described in Sect. 3.3 (*Easy and Hard online banking*, and *Easy and Hard service desk*). Cognitive support is offered *during* the exercises. Bloom’s [18] taxonomy of keywords has been used to identify all cognitively challenging elements in the

exercises, including (long) difficult words, complex scenario-specific terms, and necessary exercise steps that may not be intuitive. The coach knows when the user is having difficulty with these challenges and offers scaffolding support that ranges from ‘asking the user if they need help’ to ‘telling the user what to do’. If the user asks a question, the coach uses general-purpose utterances to answer it. In the service desk exercises, the coach can also show the user images of a Dutch passport, ID, or driver’s license. Affective support is offered *after* the exercises. The coach knows the user’s accuracy and completion time and uses this to provide motivational interviewing feedback. The coach also asks the user’s opinion on either the online banking website or the conversation partner, to encourage the user to reflect on their experience. Social support is offered *before* the exercises. The coach follows a short small talk script based on the topic of the exercise. The coach asks about the user’s experiences with and opinions on the topic of the exercise; depending on user answers, follow-up questions may be asked as well.

Interaction methods Users use a mouse and keyboard to navigate and use the online banking websites. Users can talk to the service desk conversation partner ECA and the coach ECA in natural language. For the purposes of the evaluation, the ECAs are designed to be controlled via the Wizard-of-Oz method [64]. Both ECA programmes contain a list of pre-recorded natural language utterances, which were written and recorded during prototype development: All coach utterances were voiced by one research confederate, while the conversation partner ECA voices were voiced by two other research confederates. The wizard operator controls the ECAs by selecting these utterances in a control program, causing the ECA to ‘say’ the utterance. Apart from selecting pre-recorded utterances, the wizard has no further control over the ECAs; the ECAs’ possibility space is fully described by the utterances. The wizard is not allowed to interact with participants in any other way.

Table 1 Refined VESSEL requirements baseline, from the perspective of VESSEL as an ECA coach supporting exercises

General requirements	Coach requirements	Exercise requirements
<p><i>R1. Adaptability</i> VESSEL should offer and/or support different learning styles and preferences. The focus of adaptability should be on providing the right level of difficulty (as perceived by the learner). Exercises should be difficult enough to be useful, but not so difficult that they scare low-literate learners off.</p>	<p><i>R1.1-C.</i> The coach should adapt its interaction style to the individual user’s needs, wishes, and learning goals. Coach support should ensure that exercises fall inside the Zone of Proximal Development: exercises should neither be ‘too easy’ nor ‘too difficult’.</p>	<p><i>R1.1-E.</i> The exercises should each have a specific difficulty level, tailored to particular skill training and learning goals. The total corpus of exercises should span a range of difficulty levels.</p>
<p><i>R2. Sensitivity</i> VESSEL should use non-confrontational language and content, demonstrate cultural awareness, and take existing emotional issues with regard to reading and writing and societal participation into account. The principal emotional barriers to address with sensitivity are: fear, shame and anger. Low-literate learners should feel emotionally comfortable, and experience being taken seriously.</p>	<p><i>R2.1-C.</i> The coach should always address learners calmly and kindly and avoid using phrases and broaching topics that upset low-literate learners.</p>	<p><i>R2.1-E.</i> The exercises should be as sensitive as needed to reach the intended learning goals and difficulty level.</p>
<p><i>R3. Situatedness</i> VESSEL should use learning materials and contents that are closely related to the learner’s physical environment and real-life experiences. Correctness of experience is the most important part of situatedness: the experience of training must be as close as possible to the real-life situation being trained. Learning exercises must teach low-literate learners to deal with cognitively, affectively, and socially challenging situations.</p>	<p><i>R3.1-E.</i> The exercises should use content drawn from crucial practical situations, tailored to and situated in the specific day-to-day experiences of low-literate learners.</p>	
<p><i>R4. Collaboration</i> VESSEL should have systems in place that enable, support, and foster social interaction and collaboration in learning. For low literates, it is preferable to have collaboration come from non-digital sources. If collaboration is built into the software, it must emphasize the availability of teachers and low-literate peers.</p>		
<p><i>R5. Multimodality</i> VESSEL should employ multi-modality, offering content in multiple concurrent ways. Modality use must be adapted to individual preferences and to particular exercises. Using more modalities is better than using just one.</p>	<p><i>R5.1-C.</i> The coach should combine audio ‘speech’ with visual and textual supporting material.</p>	<p><i>R5.1-E.</i> The exercises should be as multimodal as needed to reach the intended learning goals and difficulty level.</p>
<p><i>R6. Support</i> VESSEL should possess built-in support options. It is important to invoke the feeling of being supported. The right individual level of support must be found: too little support drives low-literate learners off, but too much support hampers learning and trades progress for comfort</p>	<p><i>R6.1-C.</i> The coach should use verbal scaffolding techniques to offer cognitive learning support. <i>R6.2-C</i>The coach should use motivational interviewing techniques to offer affective learning support. <i>R6.3-C</i>The coach should use small talk to offer social learning support.</p>	
<p><i>R7. Interactivity</i> VESSEL should employ real interactivity in offering content. Interactive exercises should be used to help low-literate learners practice their worst-case-scenario fears, and to learn applicable skills and gain experience.</p>	<p><i>R7.1-C.</i> The coach should interact with users actively by starting conversations and offering help. <i>R7-2-C</i> The coach should react to learner questions and demands for help.</p>	<p><i>R7.1-E.</i> The exercises should be interactive, requiring learners to use input mechanics to engage with the virtual environment in order to complete them.</p>

Table 1 (continued)

General requirements	Coach requirements	Exercise requirements
R8. <i>Gaming principles</i> VESSEL should use elements and principles of interactive gaming. Gaming principles should be used carefully, as they can be seen as childish. If gaming principles are used in the software, they should focus on evoking pride and a sense of achievement.	R8.1-C. The coach should focus on praising the learner for success over emphasizing learner failures.	Column 'General Requirements' shows the high-level requirements that were created in [88]. Columns 'Coach Requirements' and 'Exercise Requirements' show zoomed-in requirements created in Appendix A

The ECA coach has access to four groups of utterances: *Cognitive support utterances*, *affective support utterances*, *social support utterances*, and *general-purpose utterances* like “yes”, “no”, “I don’t know”, and “I did not understand you”. The wizard uses these utterances in accordance with the following rules. At the start of an exercise, social support is used. The wizard must follow the ‘small talk’ social support script as closely as possible, selecting utterances from the list of social support utterances in a pre-described order. Some of these utterances are questions that the coach asks of the user: If users answer these questions, the wizard must interpret the user’s speech and choose the correct response for the coach from a small list of possible responses. During the exercise, cognitive support is used. The wizard must interpret the user’s actions and speech to choose appropriate utterances from the list of cognitive support utterances. If users are struggling with a pre-identified challenging element, the coach should offer support about that element. In these cases, the wizard must use their own expertise to judge on a case-by-case basis which users are ‘struggling’, and which specific cognitive support utterance to use. After the exercise, affective support is used. The wizard must follow one of four motivational interviewing scripts, depending on the user’s performance (the exercise was completed with little coach support vs. the exercise was completed with a lot of coach support or not completed) and speed (the exercise was completed in under 3 min vs. the exercise was completed in 3 to 6 min or not completed). Finally, the use of general-purpose utterances is up to the wizard’s interpretation of the user’s speech and actions: This includes reacting to unanticipated user questions, prompting the user to repeat themselves if their speech was not understood, and getting the exercise ‘back on track’ as quickly as possible should unanticipated questions or disturbances occur.

The conversation partner ECA has access to two groups of utterances. A scenario script contains all utterances, in the correct order, to hold the exercise conversation. The wizard must follow the scenario script perfectly when the exercise is conducted. Here, too, some of the ECA’s utterances are questions; the wizard must interpret user answers to these questions to select the correct follow-up utterance. A second list contains general-purpose utterances, similar to the coach’s (but recorded in the conversation partner ECA’s voices).

Appearance Exercise appearances are shown in Sect. 3.3 (see Figs. 3 and 4). The coach ECA’s visual appearance was based on the same pre-study used for the conversation partner ECAs [26]. Participants were asked to imagine the twelve ECA characters as digital coaches and order them from most to least preferred. While not unanimous, one particular ECA was ranked in the top spot more than any other. We selected this ECA as the coach (see Fig. 5). Like the conversation partner ECAs, the coach ECA has one facial



Fig. 5 VESSEL Coach ECA (top right corner) and supporting material for the Easy online banking exercise. Text is in Dutch. From top to bottom, lines read: ‘The exercise is: Transferring money using online banking. To whom: Mister Jansen. How much money: 10 euro. Account number: NL POST 1200 1111 00’

expression, opens and closes its mouth while talking on a single animation cycle regardless of the audio being played, does not gesture or use body language, and animates in a simple ‘swaying its head back and forth’ animation loop. A grey background is used.

6 Evaluation: methods

6.1 Experimental design

An experiment was carried out to evaluate the learning effectiveness impact of our VESSEL prototype coach, in terms of the six claims that were presented as use case post-conditions (see Appendix B). In the SCE method, specification claims serve as evaluation hypotheses. This results in the following six hypotheses:

Learning Experience

- *H1: Cognitive Experience (Performance)* The coach leads to a shorter exercise completion time, and higher perceived performance.
- *H2: Affective Experience (Positive Affect)* The coach leads to more positive affective states during and after the exercise.
- *H3: Social Experience (Engagement)* The coach increases the amount of user-system interaction and results in learners viewing VESSEL as more helpful and easy to learn with.

Learning Outcomes

- *H4: Cognitive Outcomes (Success)* The coach leads to a higher exercise completion rate.
- *H5: Affective Outcomes (Self-Efficacy)* The coach leads to higher self-efficacy.
- *H6: Social Outcomes (Retention)* The coach leads to a higher motivation to continue learning.

To test these hypotheses, a mixed-method repeated-measures within-subjects experiment was designed. The study’s main independent variable was *Coach Presence*. This variable had two levels: *With Coach*, and *Without Coach*. Participants were invited to work with the prototype in two consecutive sessions (one week apart): One session in which they tested the complete prototype, including all exercises and the ECA coach (the ‘With Coach’ session), and one session in which they tested a version of the prototype that only included the four exercises, but not the coach (the ‘Without Coach’ session). In the With Coach session, participants completed all four exercises (*Easy online banking*, *Hard online banking*, *Easy service desk*, and *Hard service desk*) with support from the coach. In the Without Coach session, participants completed the same exercises without coach support. All participants participated in both sessions. Session order was counterbalanced: 50% of participants did the With Coach session the first week and the Without Coach session the second week, and 50% of participants did the opposite. Exercise order was partially counterbalanced: Each participant was offered the four exercises according to one of four pre-determined orderings. These orderings were counterbalanced across participants, but kept the same per participant in both the With Coach and Without Coach sessions.

6.2 Measures

Twenty-dependent variables were measured: Eighteen variables were self-report measures, obtained using two questionnaires (see Sect. 6.4), and two variables were objective performance metrics. Table 2 describes the variables. Additionally, semi-structured interviews were used to gain qualitative insight into the proceedings and the participants’ experiences with the VESSEL prototype, with the following questions:

- How did you like the session? Do you think it went well, or poorly?
- What went well for you? What went poorly for you? What did you think was the cause?
- What parts of the session did you enjoy? And what parts of the session did you dislike?
- What would you change about the exercises you just did?

One question was only asked in the With Coach session:

- What do you think about the coach? Has the coach helped you? Was it nice to have the coach around, or annoying?

Additionally, the following questions were only asked after the second session:

- Did you notice any differences between the two sessions? What differences did you see?
- Which of the two sessions did you like best? And why?

6.3 Participants

Twelve low-literate users participated in the entire experiment. Kurvers et al.'s [56] five language learner profiles were used to select suitable participants; these profiles divide low-literate first-language learners (L1) and second-language learners (L2) into categories based on their language comprehension skills and their learning ability. Only learners that matched profiles 2 (L1 and L2 learners with no particular strengths or weaknesses, considered 'average low-literate learners'), 3 (typical L2 learners, particularly struggling with vocabulary and with speaking and understanding spoken Dutch), and 4 (low-skilled L1 learners, with decent speaking skills but serious difficulties with reading and writing) were invited to participate, as these learners can realistically benefit from computer-supported learning. Learners at the extreme ends of the low-literacy spectrum (profiles 1, relatively high-skilled and self-directed L1 and L2 learners, and 5, L1 and L2 learners with serious learning difficulties and very limited educational backgrounds) are expected, respectively, to be too skilled to benefit from our support, and too low-skilled to be able to use our prototype in the first place. Participants were recruited in several language classes throughout the Netherlands; teachers in these classes used the profiles to select and invite suitable learners to participate. Six men and six women participated, with ages ranging from 30 to 63 ($M=48.2$, $SD=10.5$). Two of the participants were natively fluent in Dutch; the other ten participants identified as 'somewhat fluent'. Other (native) languages spoken by the participants included: Arabic, Bosnian, Edo, English, French, Somali, Spanish, and Turkish. Four participants reported having prior experience with online banking, and all twelve participants had prior experience with service desk conversations. Of the latter, seven participants specifically had experience with passport recovery. There was no overlap between these participants, and the participants for the pre-study [26].

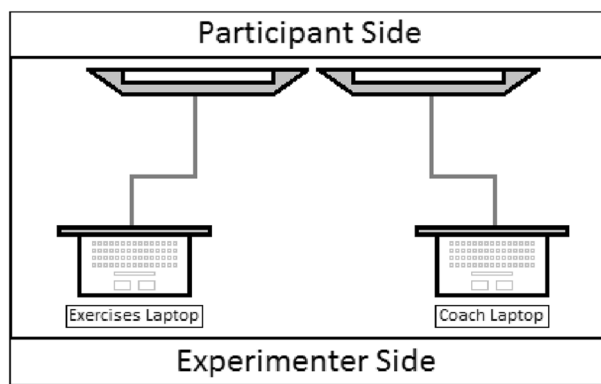


Fig. 6 Schematic overview of experimental setup: 2 laptops (lower figures) connected to 2 monitors (upper figures). Monitors are placed and angled such that participants could not see the laptops and the experimenters while seated

6.4 Materials

The experimental setup consisted of two laptops connected to two additional monitors (Fig. 6). The laptops, on one side of the table, were used by the experimenters. The monitors, on the other sides of the table, were used by the participants. The laptop and monitor on the right were used to run and control the coach. The laptop and monitor on the left were used to run and control the exercise environment. On the participant side, a mouse and keyboard (plugged into the left laptop) were provided for the online banking exercises.

Three questionnaires were used. Two questionnaires measured the eighteen self-report variables (see Table 2): These were called the 'exercise' questionnaire (EQ), and the 'session' questionnaire (SQ). Answers were given on a five-point bipolar Likert scale, using greyscale answer bars (Fig. 7). Participants were told to mark one of the five boxes per question. Bars were labelled 'Nee' (No) and 'Ja' (Yes) at the left and right extremes. Question SQ.1 was included as a practice question to allow low-literate participants to 'get used to' the answer schema and was not included in later analysis. A third 'demographic' questionnaire measured participant age, sex, schooling history, time spent in the Netherlands, known languages ('fluent' and 'somewhat fluent'), and prior experiences with online banking and city hall service desk situations. For objective measures, exercise completion time was measured with a stopwatch, and exercise completion was tallied by hand. Finally, an audio recorder was used to record the experimental proceedings and the end-of-session interviews.

6.5 Procedure

The first session started with a general introduction, informed consent forms, and the demographic

Table 2 Overview of quantitative measures

Variable	Hypothesis	Description
<i>Subjective measures: exercise questionnaire (EQ)</i>		
EQ.1 Perceived performance (exercise)	H1	“I have done the exercise about (<i>online banking / talking to people at a service desk</i>) well.”
EQ.2 Difficulty (exercise)	H5	“I found the exercise to be difficult.”
EQ.3 Self-efficacy (exercise)	H5	“I am now better at (<i>online banking / talking to people at a service desk</i>).”
EQ.4 Positive affect (exercise)	H2	“I am happy with how I did the exercise.”
EQ.5 Computer support	H3	“The computer helps me to do the exercise well.”
<i>Subjective measures: session questionnaire (SQ)</i>		
SQ.1 Positive affect (language class)	(<i>practice question</i>)	“I like coming to language class.”
SQ.2 Self-efficacy (online banking)	H5	“I am good at online banking.”
SQ.3 Self-efficacy (spoken Dutch)	H5	“I am good at understanding spoken Dutch.”
SQ.4 Self-efficacy (service desk)	H5	“I am good at talking to people behind a service desk.”
SQ.5 Self-efficacy (written Dutch)	H5	“I am good at reading written Dutch.”
SQ.6 Self-efficacy (computer use)	H5	“I am good at working with a computer.”
SQ.7 Computer usefulness	H3	“A computer helps me learn.”
SQ.8 Positive affect (online banking)	H2	“I enjoy online banking.”
SQ.9 Positive affect (service desk)	H2	“I enjoy talking to people behind a service desk.”
SQ.10 Difficulty (online banking)	H5	“I find online banking to be difficult.”
SQ.11 Difficulty (service desk)	H5	“I find it difficult to talk to people behind a service desk.”
SQ.12 Desire to improve (online banking)	H6	“I want to get better at online banking.”
SQ.13 Desire to improve (service desk)	H6	“I want to get better at talking to people behind a service desk.”
<i>Objective measures: measured per exercise</i>		
DM.1 Completion level (exercise)	H4	Binary value: Whether or not participant completed the exercise.
DM.2 Completion time (exercise)	H1	Exercise completion time in seconds.

Includes measure source (EQ, SQ, or direct measurement), applicable hypothesis, and question wording or measure description

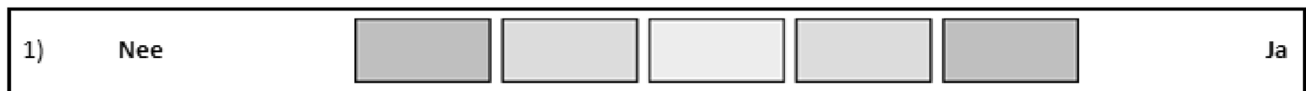


Fig. 7 Example answer bar for the short and SQs (‘Nee’ means ‘No’, ‘Ja’ means ‘Yes’)

questionnaire. Next, the first SQ was administered. Experimental proceedings diverged after that, based on experimental condition. In the Without Coach condition, researchers explained the general experiment flow. Participants were introduced to their first exercise and shown the instruction material. Participants were told to complete the exercise alone, without outside help, within the 6-minute time limit (which they could not see). After that time, or as soon as participants were finished, an EQ was administered. Proceeding from there, the remaining exercises were carried out the same way. In the With Coach condition, researchers instead introduced the coach. The coach (controlled always by the same experimenter) introduced

itself to the user (with the name ‘Anna’) and explained the general experiment flow. The coach introduced the first exercise and the instruction material and offered social learning support. Participants were told to complete the exercise, with the help of the coach, within the 6-minute time limit. During the exercise, the coach provided cognitive learning support when needed. After the time limit, or as soon as participants completed the exercise, the coach offered affective learning support. Researchers then administered an EQ. The remaining exercises were carried out the same way. After the end of the fourth exercise, conditions converged. The researchers administered a second SQ. Then, a semi-structured ending interview was conducted, using the questions presented above. And finally, participants were debriefed.

In the second week of experimental sessions, each participant completed the opposite set of exercises, swapping the With Coach and Without Coach conditions. Otherwise, the same procedure and exercises from week 1 were used. The week 2 ending interview used the same questions as week 1, but included the questions about the perceived differences between the two conditions. Finally, at the end of week 2, participants were fully debriefed and rewarded for their participation.

7 Evaluation: results

Three sets of analyses were carried out. First, a repeated-measures general linear model (GLM) analysis was conducted on the EQ data. Second, a factor analysis was used to condense the data of the SQs into several factors; another repeated-measures GLM analysis was conducted on these results, as well as a paired-samples T-test. Third, a final repeated-measures GLM analysis was used to analyze the performance results of the online banking exercises. Finally, qualitative observations were made by the researchers, both live during the experiment and by listening to the audio recordings afterwards.

Prior to analysis, questionnaire reliabilities were investigated. The EQ had an average reliability of $\alpha = .845$. Beside question 1 (the ‘practice question’), question 13 was also dropped from the SQ as it showed scattered answers and low reliability (based on general descriptives and Cronbach’s alpha). The complex wording of this question seems to have led to confusion and misunderstanding. The remaining eleven questions show an average reliability of $\alpha = .600$.

7.1 Exercise questionnaire analysis

A 2-by-2-by-2 repeated-measures GLM analysis was done with the EQ data. Three GLM factors were chosen. The *Coach* factor had two levels: ‘With Coach’ and ‘Without Coach’. The *Scenario* factor had two levels: ‘online banking’ and ‘service desk’. The *Difficulty* factor had two levels: ‘Easy’ and ‘Hard’. The five questions of the EQ were all treated independently: They were designed to measure entirely separate concepts, and Pearson correlation analysis showed no significant correlations. All main effects and all interaction effects were tested. Table 3 shows means and standard deviations of the five questions for each of the eight measurement moments. Table 4 shows the analysis results.

The following significant results were found:

- *Coach* Results showed that perceived performance was higher, positive affect was higher, and perceived computer support was higher for With Coach compared to Without Coach
- *Scenario* Results showed higher perceived performance, higher self-efficacy, and higher positive affect for service desk compared to online banking. Online banking showed higher experienced difficulty
- *Difficulty* Results showed higher perceived performance, higher self-efficacy, and higher positive affect for Easy compared to Hard. Hard showed higher experienced difficulty
- *Coach*Scenario* Two sets of effects were found. For online banking only, the With Coach condition showed increased perceived performance, positive affect, and perceived computer support compared to Without Coach. This was not seen for service desk. Furthermore, With Coach showed lower experienced difficulty for online

Table 3 EQ means (*standard deviations*)

	With coach				Without coach			
	Online banking		Service desk		Online banking		Service desk	
	Easy	Hard	Easy	Hard	Easy	Hard	Easy	Hard
EQ.1 Perceived performance (exercise)	1.25	.00	1.50	.75	.58	−1.50	1.33	.75
EQ.2 Experienced difficulty (exercise)	.75	1.54	.52	1.54	1.68	1.17	.98	1.14
EQ.3 Self-efficacy (exercise)	−1.25	.92	−1.50	−.75	−.42	1.58	−1.67	−1.33
EQ.4 Positive affect (exercise)	.97	1.00	.90	1.48	1.68	.90	.89	1.15
EQ.5 Computer Support	.50	.33	1.42	1.08	.75	−.67	1.08	.67
	1.31	1.30	.67	1.08	1.36	1.61	1.08	1.44
	1.50	.75	1.58	1.41	.92	−.50	1.67	1.42
	.80	1.50	.51	.67	1.62	1.57	.65	.67
	1.50	1.58	1.08	1.17	1.00	−.50	.92	1.00
	.52	.67	1.56	1.11	1.54	1.73	1.51	1.48

Mean scores range from [−2,2]

Table 4 Significant results of EQ repeated-measures GLM analysis

	Coach	Scenario	Difficulty	Coach* Scenario	Coach* Difficulty	Scenario* Difficulty	Coach* Scenario* Difficulty
EQ.1 Perceived Performance (exercise)	$F = 15.40$ $p = .00$ $\beta = .95$	$F = 14.67$ $p = .00$ $\beta = .94$	$F = 40.68$ $p = .00$ $\beta = 1.00$	$F = 8.25$ $p = .02$ $\beta = .74$		$F = 12.00$ $p = .01$ $\beta = .83$	
EQ.2 Experienced Difficulty (exercise)		$F = 39.11$ $p = .00$ $\beta = 1.00$	$F = 54.92$ $p = .00$ $\beta = 1.00$	$F = 8.76$ $p = .01$ $\beta = .77$		$F = 19.95$ $p = .00$ $\beta = .98$	
EQ.3 Self-efficacy (exercise)		$F = 7.23$ $p = .02$ $\beta = .69$	$F = 16.84$ $p = .00$ $\beta = .96$				$F = 5.67$ $p = .04$ $\beta = .58$
EQ.4 Positive Affect (exercise)	$F = 11.80$ $p = .01$ $\beta = .88$	$F = 10.78$ $p = .01$ $\beta = .85$	$F = 14.00$ $p = .00$ $\beta = .93$	$F = 11.30$ $p = .01$ $\beta = .86$		$F = 6.10$ $p = .03$ $\beta = .62$	
EQ.5 Computer Support	$F = 6.42$ $p = .03$ $\beta = .64$			$F = 15.10$ $p = .00$ $\beta = .94$	$F = 7.05$ $p = .02$ $\beta = .68$	$F = 4.95$ $p = .05$ $\beta = .53$	$F = 6.50$ $p = .03$ $\beta = .64$

For each question and each factor / group of factors, *F*-value (*F*), significance (*p*), and observed power (β) are given if $p < 0.05$

banking, but higher experienced difficulty for service desk; Without Coach did not show this

- *Coach*Difficulty* Results showed that With Coach raised self-efficacy in the Hard exercises compared to Without Coach. No similar effect was seen for the Easy exercises.
- *Scenario*Difficulty* Results showed that for online banking, the Hard exercise resulted in lower perceived performance, higher experienced difficulty, lower positive affect, and lower perceived computer support, compared to the Easy exercise. No similar effects were seen for service desk
- *Coach*Scenario*Difficulty* Two effects were found. In the *Hard online banking* exercise, With Coach showed

higher self-efficacy than Without Coach; in the *Easy online banking* exercise, no difference was found. Perceived computer support was much higher for the *Easy online banking* exercise than for the *Hard online banking* exercise, although both dropped significantly in the Without Coach condition compared to With Coach. In both cases, no effects were seen for service desk

Tests for between-subjects effects showed no significant results for age, gender, experience with online banking/ service desk, and exercise counterbalancing order.

Table 5 Factor loadings for the 11 questions used in the factor analysis

Item	Factor			
	1	2	3	4
SQ.2 I am good at online banking	.874			
SQ.5 I am good at reading written Dutch	.858			
SQ.6 I am good at working with a computer	.855			
SQ.10 I find online banking to be difficult	.761			
SQ.3 I am good at understanding spoken Dutch		.587		
SQ.4 I am good at talking to people behind a service desk		.873		
SQ.9 I enjoy talking to people behind a service desk		.875		
SQ.11 I find it difficult to talk to people behind a service desk		.821		
SQ.7 I computer helps me learn			.740	
SQ.12 I want to get better at online banking			.814	
SQ.8 I enjoy online banking				.841

Only factor loadings of .500 and higher are shown

7.2 Session questionnaire analysis

Three analysis steps were used for the SQ. First, factor analysis was used to effect data reduction: Pearson correlation analysis showed several potentially significant correlations. An exploratory factor analysis was conducted, using principal component analysis for extraction and Varimax Rotation (with Kaiser Normalization) for rotation. Both eigenvalues and scree plots suggested a solution with four factors. Table 5 shows the factor loadings for this solution.

Based on the factor loadings shown in Table 5, a two-factor solution was decided on. Factor 1, ‘Information Skills’, contained questions 2, 5, 6, and 10, with a reliability of $\alpha=.86$. Factor 2, ‘Communication Skills’, contained questions 3, 4, 9, and 11, with a reliability of $\alpha=.82$. While questions 7 and 12 seemed to form a third factor, the reliability of this factor was only $\alpha=.41$; these questions were kept as separate items instead. Question 8 was also kept as a separate item.

Second, a 2-by-2 repeated-measures GLM analysis was conducted on the two factors and three independent questions. Because the SQ was only administered at the start and end of each experimental session, only two GLM factors were chosen. The *Coach* factor had levels corresponding to the coach’s presence or absence, ‘With Coach’ and ‘Without Coach’, and the *Time* factor had levels corresponding to the SQ measurement moment, ‘Pre-Session’ (the questionnaire was administered before a session) and ‘Post-Session’. (The questionnaire was administered after the end of a session.) All main effects and interaction effects were tested. Only one significant result was observed: Participant information skill was higher for Post-Session than for Pre-Session ($F=5.474$, $p=.039$). Tests for between-subjects effects showed no significant results for age, gender, experience with online banking/service desk, and exercise counterbalancing

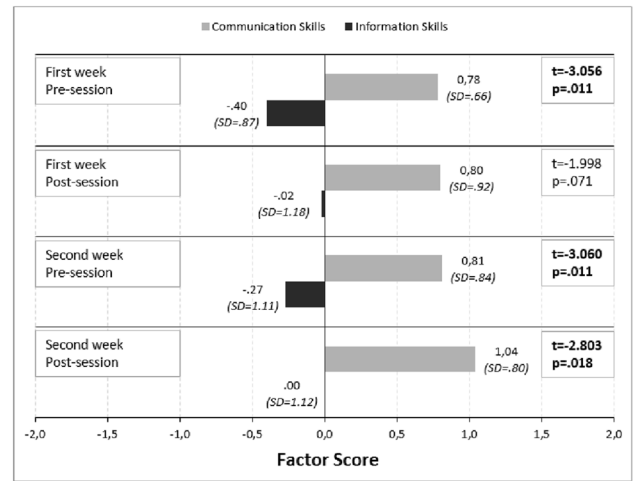


Fig. 8 Means for ‘Information Skills’ and ‘Communication Skills’ factors, across the four measurement moments. Values next to bars represent mean (standard deviation). Boxes on the right show the test statistics of the paired-samples T-test that compared means for Information Skills and Communication skills in that measurement moment; bold text indicates significant results

order. Table 6 shows means and standard deviations for the factors and questions.

Third, a paired-samples T-test was conducted to compare ‘Information Skills’ and ‘Communication Skills’ means in the first and second week. These means are different from the means in Table 6: 50% of participants did With Coach sessions in the first week, and 50% did Without Coach in the first week. First week/second week means were compared for the four SQ measurement moments (before and after each session). Results are shown in Fig. 8. Before the first experimental session and before and after the second session, ‘Communication Skills’ were significantly higher than ‘Information Skills’.

Table 6 SQ data means (standard deviations)

	With coach		Without coach	
	Pre-session	Post-session	Pre-session	Post-session
Factor 1:	-.27	.10	-.40	-.13
Information skills	1.07	1.20	.91	1.08
Factor 2:	.73	.88	.86	.97
Communication skills	.81	.88	.69	.86
SQ.7 A computer	.92	1.41	1.33	1.33
Helps me learn	.79	.67	.78	.78
SQ.8 I enjoy online banking	-.08	.67	.17	.42
	1.31	1.30	1.53	1.62
SQ.12 I want to get	1.41	1.83	1.08	1.58
Better at online banking	1.24	.39	1.38	1.16

Mean scores range from [-2,2]

7.3 Performance metrics analysis

A 2-by-2 repeated-measures GLM analysis was done with participant completion time and completion rate. Only data from the online banking exercises were used for this analysis: Data from the service desk exercises did not show enough variance, as completion rates were 100% for both exercises and completion times were strongly homogeneous. Two GLM factors were chosen: The *Coach* factor had levels: ‘With Coach’ and ‘Without Coach’, and the *Difficulty* factor had levels: ‘Easy’ and ‘Hard’. All main effects and interaction effects were tested. Significant results were only seen for the Difficulty factor: Exercise completion time was higher ($F= 13.035, p=.006$), and exercise completion rate was lower ($F=22.559, p=.001$) for Hard compared to Easy. Tests for between-subjects effects showed no significant results for age, gender, experience with online banking / service desk, and exercise counterbalancing order. Table 7 shows means and standard deviations.

7.4 Observations

Experimenters observed that the coach seemed to work as intended, particularly for the online banking exercises. In Without Coach sessions, participants often seemed to quietly struggle with completing the exercise; no participants tried to talk to the computer system, and only some participants tried to get researcher help. However, in With Coach sessions, almost all participants interacted with the coach in some way and benefited from its help. Broad personal differences were observed in the degree to which this happened. Participants who spoke with the researchers a lot during the introduction to the experiment spoke to the coach in the same way they would talk to a human actor, including attributing personality traits to ‘her’ and asking complex questions (e.g. “*This bill I have to pay seems way too high. Coach, what do you think?*”). Participants who spoke less to the researchers commonly spoke less to the coach as well, generally restricting themselves to answering coach questions and asking for direct instructions (e.g. “*Coach, how do I pay a bill?*”). However,

these participants were still seen acting on the coach’s advice. Significantly less coach-participant interaction was seen during the service desk exercises. Participants asked for help less often, and in fact seemed to get stuck less often. Interestingly, whenever help was needed, participants more often asked the service desk ECA directly. With the focus on the conversation partner, participants seemed to overlook the coach’s presence. One participant echoed this, saying that (paraphrased) “... *I kind of forgot she was there.*” Participants spoke to the service desk ECA the same way they spoke to the coach, i.e. some participants really engaged with her, while others only answered direct questions. Interestingly, this most often happened in situations where the scenario was inaccurate or incomplete compared to real life: For instance, many participants asked if they would be required to ‘bring passport photos next time’, something that we had not incorporated in the exercise. Participants would use their own experience and expertise with these scenarios to catch these inaccuracies, and then press the conversation partner for clarification.

Some unexpected technical difficulties occurred during the experiment. Both the coach and the conversation partner programmes showed an unexplained, variable time delay when speaking, ranging from two seconds to twenty. From interviews, it seems participants perceived this as ‘the coach just being very quiet’. But for the experimenters, this made it hard to use the right support at the right time. Particularly in situations where participants asked questions and then quickly moved on, this was a problem: The coach would either be stuck using a now-irrelevant speech utterance, confusing the participant, or it would have to be muted for the duration, causing a strange visual effect (the coach soundlessly ‘talking’) that some participants noticed. In either case, no further support would be possible for a while.

Post-test interviews showed that most participants accepted the Wizard-of-Oz illusion quite readily. Participants did not notice the behind-the-scenes technical difficulties, and even the aforementioned ‘soundless talking’ was usually mentioned as an oddity, not as something that stood out. When asked about the coach, participant response was almost universally positive. This seemed inversely correlated with ‘participant skill’: Participants who completed the exercises easily and quickly were more often ambivalent or negative about the coach, while participants who required a lot of help to complete exercises were very happy with the coach’s help. Participants were positive about the entire VESSEL prototype: Many mentioned that they enjoyed this way of learning and doing exercises and expressed hope that they would be able to ‘do something like this at home’ soon. The interviews also showed that many participants saw the online banking exercises as much more difficult than the service desk exercises. Particularly, the *Hard online banking* exercise was considered very challenging, and almost impossible to complete (within the time limit) without the coach.

Table 7 Performance metrics means (*standard deviations*) for online banking exercises

	With coach		Without coach	
	Easy	Hard	Easy	Hard
DM.1 Completion Time (s)	280.2 <i>128.3</i>	364.8 <i>113.4</i>	209.3 <i>107.2</i>	347.9 <i>77.8</i>
DM.2 Completion Rate	.83 .39	.50 .52	.75 .45	.17 .39

Completion time ranges from [0–600] (in seconds). Completion rate ranges from [0–1]

The coach's support was much appreciated here. In contrast, both service desk exercises were seen as easy. Participants did sometimes notice differences in politeness between the two conversation partners, but otherwise seemed to consider the exercises equivalent; this was not the case with the online banking exercises, which were more clearly 'easy' and 'difficult'.

8 Conclusions and discussion

8.1 Findings

Building on earlier design work for the envisioned system VESSEL [88, 89], this study has designed, developed, and evaluated VESSEL as a virtual learning environment, wherein societal participation exercises are supported by an ECA coach. The foundation of data was updated with situated interactive exercises, the literature on cognitive, affective, and social learning support, and the benefits of ECA coaching. The specification requirements were refined to reflect VESSEL as an ECA coach-supported exercise environment (see Table 1). Use cases were derived and used to design and develop a functional VESSEL prototype. This prototype was tested with low-literate end users in order to evaluate the claims of learning effectiveness underlying the ECA coach.

The study's first research question was: "*How can we create an ECA coach that provides effective cognitive, affective, and social learning support to low-literate learners doing situated interactive exercises in a virtual learning environment?*" Sub-question 1a, "*In what ways can an ECA coach provide cognitive, affective, and social learning support?*", is answered in Sect. 3. The coach should offer cognitive support in the form of scaffolding, affective support in the form of motivational interviewing, and social support in the form of small talk. Sub-question 1b, "*Which functionalities, interaction methods, and appearances should an ECA coach have to effectively provide this learning support in a virtual learning environment?*" is answered in Sects. 4 and 5. The coach should provide learning support that is adapted to the individual learner, to help them complete exercises. The coach should interact with learners in the form of pre-recorded utterances, complemented with visual materials when necessary; and the coach's appearance should align with user expectations of the role of a 'digital coach'. This outcome seems true across all participating learners, regardless of age, sex, or ethnicity. Expectations for the pre-study [26] were that participants would prefer ECAs that were similar to them in gender and ethnicity, as humans can experience similarity attraction to ECAs just as to humans [70, 72]. But instead, all ECAs in the prototype were valued on

matching the (visual) stereotype of their role. Participants chose the coach depicted in Fig. 5 because 'she looked friendly and approachable', and the *Easy service desk* conversation partner depicted in Fig. 4 because 'she looked like she belonged there, like she would know what was happening'. Participants would actually dislike similar-looking ECAs, saying that (paraphrased) "*if this person is like me, also low literate, then they won't be able to help me*". This clashes with expectations that user-ECA similarity attraction would be high [70], but does confirm that stereotype-reinforcing appearances can have a strong impact [4]; our results seem to suggest a 'job-appropriate clothing' stereotype rather than Angeli & Brahmam's [4] sex and gender stereotypes, though it should be mentioned that our 'most positive' ECAs were both read as female, while our 'most negative' ECA was read as male. Our most-popular ECAs were also generally the more conventionally attractive ones, mirroring results by Khan and de Angeli [49] and Nass et al. [72]. In this study's post-test interviews, our participants (no overlap with the pre-study participants) primarily reported that they judged the ECA characters on how well they fit the scenario: The coach and the *Easy service desk* conversation partner were liked and valued, while the *Hard service desk* conversation partner was disliked.

The study's second research question was: "*Does this support-providing ECA coach increase learning effectiveness for low-literate learners working with VESSEL, compared to low-literate learners working with VESSEL but not receiving coach support?*" Six hypotheses were derived, based on six claims of learning effectiveness: Cognitive, affective, and social learning experience and cognitive, affective, and social learning outcomes. Using the results from Sect. 7, these hypotheses resolve in the following ways:

Learning Experience

- *H1: Cognitive Experience (Performance)* This hypothesis is partially supported. Self-reported performance increased in the presence of the coach. However, completion time did not. Users experienced that they were doing better in the presence of the coach, but were not actually any faster
- *H2: Affective Experience (Positive Affect)* This hypothesis is supported. User positive affect significantly increased in the presence of the coach
- *H3: Social Experience (Engagement)* This hypothesis is supported. Users reported feeling 'supported by the computer' significantly more when the coach was present. Users were also observed to interact with the system much more when the coach was present: Users actually talked to the coach, with some interactions going beyond the exercise topics

Learning Outcomes

- *H4: Cognitive Outcomes (Success)* This hypothesis is not supported. No significant main effect of coach presence was found for exercise completion rate
- *H5: Affective Outcomes (Self-Efficacy)* This hypothesis is partially supported. No significant main effect of coach presence on any self-efficacy measure was found. But an interaction effect shows that the coach significantly raised online banking self-efficacy only after the *Hard online banking* exercise
- *H6: Social Outcomes (Retention)* This hypothesis is not supported. After factor analysis, only SQ question 12 measured this hypothesis. No significant main effect of coach presence was found

The ECA coach created in this study, designed to provide cognitive, affective, and social learning support meaningfully integrated into four online banking and service desk exercises, has significantly increased several aspects of the learning effectiveness of VESSEL. The hypothesis that working with the ECA coach would improve the learning experience is fully supported in hypotheses H2 and H3, and partially supported in H1. The hypothesis that working with the coach would improve learning outcomes is only partially supported in hypothesis H5. This seems to indicate that the coach particularly influenced participants' subjective experience of working with VESSEL: Participants were more positive and more engaged, felt like they performed better and showed a higher self-efficacy regarding online banking. A similar increase in social engagement and self-efficacy was found by Lane et al.'s [57] ECA coach, and similar improvements in the affective experience are reported by Lester et al.'s *persona effect* study [59], Lane et al.'s computer science education ECA coach [57], Bercht & Vicari's pedagogical support agent [9], and Shaw et al.'s embodied situated support agent [93]. However, objective measures of performance and success (exercise completion rate and completion time) were not influenced. This result goes counter to other studies that show that ECA coaches can influence objective learning outcomes such as learner behaviour (e.g. increasing rate of physical exercise and changing diet [16], and increasing learner involvement in dialogue with a learning agent [86]) and user-system satisfaction [14]. This discrepancy bears further investigation. It is possible that our skew towards subjective (self-reported) findings is a result of the mostly subjective set of measurements. Future studies should investigate if other objective performance measures (such as number of mistakes made, or amount of coach support needed) reveal more digital coach effects, or if the influence of the VESSEL coach as described in this work is mostly subjective.

8.2 Limitations

As this prototype was designed to be a proof-of-concept first design, a number of unexpected shortcomings were encountered during the experiments. These can be related to the functionality, interaction methods, and appearance of the exercises and the coach. A significant issue with exercise functionality was that difficulty levels of the exercises did not come out as balanced as designed. In analysis, both of the *Easy (online banking/service desk)* and both of the *Hard* exercises were treated as equivalent in difficulty level (as intended). However, the *Hard online banking* exercise was significantly more difficult than any other. This can be seen in the main effects and interaction effects for the 'Scenario' and 'Difficulty' factors: All main effects for either factor are always accompanied by either a 'Scenario*Difficulty' interaction effect, or a 'Coach*Scenario *Difficulty' one that shows strong differences between the *Hard online banking* exercise and the other three exercises. Additionally, in the post-experiment interviews, many participants reported seeing the *Hard online banking* exercise as an outlier. Almost no differences were seen between the two service desk exercises. Following up on the 'Coach*Scenario' and 'Coach*Scenario*Difficulty' interaction effects seems to suggest that all coach-related main findings only apply to 'difficult information skill exercises', or maybe only to 'difficult online banking'. Consequently, result generalizability suffers. This can be seen as a failure to adhere to requirement R1.1-E (exercise adaptability). Difficulty levels were not properly calibrated. For the online banking exercises, difficulty was intended to come from complexity and information density differences. But these differences were much stronger than expected. For the service desk exercises, difficulty was intended to come from sensitivity and politeness differences. However, strict adherence to R2.1-E (exercise sensitivity) meant that these exercises were not significantly different in practice. Additionally, participant communication skill was significantly higher than information skill, throughout the experiment. All participants reported having prior experience with 'service desk conversations', and over half of all participants had explicit experience with 'passport application' conversations. This likely lowered the experienced difficulty for these exercises. Both of these issues highlight the importance of user involvement in all steps of the design process, particularly when designing for demographics with particular needs: Pre-testing the exercises with low-literate users would have revealed the low impact of the politeness manipulation and the users' pre-existing knowledge of and focus on the exercise domain, allowing more careful calibration to take place. This stands as a lesson for future work.

service desk exercise interaction methods showed two more shortcomings. First, because participants held a natural

dialogue with the conversation partner, it turned out to be unexpectedly difficult for the coach to provide support without interfering in the conversation. To provide support, the service desk conversation would have to be stopped, creating unrealistic pauses (in scenario context). Additionally, participants reported in the interviews that switching attention between the conversation partner and the coach felt strange and took effort. Participants would direct their questions at the conversation partner instead of the coach, and (in some cases) forget about the coach entirely. While the single ECA coach in the online banking exercises has worked, having multiple (talking) ECAs in a single exercise may require more careful design; collecting all required functionality in a single ECA seems like the optimal solution (and one we intend to study in later work), but if this is not possible, user-centred design and testing could ensure that the different ECAs are actually perceived as uniquely meaningful. Second, measures of exercise completion rate and completion time were useless for the service desk exercises: Regardless of difficulty level or coach presence, exercise completion rate was 100%, and completion times showed very little variance. Again, this can primarily be blamed on high participant communication skills and experience with the scenario. Adding to this, the fact that the exercise was a conversation gave it a clear, easy-to-understand structure that the online banking exercises did not have. On the online banking websites, participants could get lost and lose time, while during the conversations, the conversation partner guided the participant with directed questions. This may have disincentivized ‘exploratory’ behaviour: Participants felt they had to follow suit in the conversation, instead of (for example) asking about unfamiliar words. The combination of prior experience and a strong guided structure meant that all participants completed the conversation in close to minimum possible time. For this type of exercise, ‘completion time’ may not be a valuable performance metric.

The most significant issue with the ECA coach was the informal nature of the Wizard-of-Oz control rules. Clear behavioural rules are important for the success of the Wizard-of-Oz-method [64]. During the small talk and motivational interviewing sections, there was a flow structure based (partially) on measurable objectives and keywords. However, particularly during the scaffolding section, the provided support was highly dependent on the wizard’s appraisal of the situation. This led to two uncertainties. Functionality-wise, it was unclear what support the coach should give at any given time and for any given problem, which can be cast as a failure to adhere to requirement R1.1-C (coach adaptability). Regarding interactions, it was unclear how much of the participants’ utterances the coach (represented by the wizard) was supposed to understand. Due to lack of clear rules, the wizard has probably responded to more participant utterances and behaviours in their interaction than an automated

ECA could have done. A human wizard can understand participant questions, perceive and read participant nonverbal cues such as body language, and analyze their emotional state. A human wizard can also apply their own reasoning to understand what a participant is ‘trying to do’, and direct support accordingly. This makes the found effects uncertain. Would a fully autonomous digital coach, with limited interaction possibilities, still have the same effects for low-literate participants? Veletsianos & Miller [104] emphasize the importance of a social, human-like experience for users working with pedagogical agents, suggesting that more machine-like interaction might not have the same positive effects. Future work should investigate ways of structuring and formalizing the coach’s control rules, regarding both support functionality and speech recognition (taking into account the opportunities afforded for the latter by state-of-the-art technology), in order to increase accuracy and study the VESSEL concept as envisioned.

One issue shared by the coach and the service desk exercises was the low graphical fidelity of the ECAs. All ECAs had a low-fidelity, somewhat unrealistic appearance, and only one facial expression. The question of whether human ECAs should be ‘naturalistic’ (as human-looking as possible) or ‘stylized’ (non-realistic and exaggerated) has no clear answer: Haake & Gulz [42] collect and discuss arguments for both approaches and conclude that the ‘right’ answer in any context depends on the agent’s intended goals and motives. While the VESSEL ECAs were accepted as social actors, it is possible that more naturalistic appearances would have resulted in stronger emotional and social bonding: Perhaps the coach’s emotional support, or the intended politeness of the *Easy service desk* conversation partner and intended rudeness of the *Hard service desk* conversation partner, would have shown stronger effects. This stands as a direction for future study.

Finally, two important oversights in the experimental design relate to the participants. First, the relatively low total number of participants almost definitely influenced result significance and power. Second, the relative lack of first-language learner (L1) participants may have made it impossible to find differences between these participants and second-language learner (L2) participants. It is well-documented that low-literate first- and second-language learners encounter different problems in learning and participation [43, 55, 56, 79]. Since we could only find two L1 participants for our evaluation, we cannot say if the two groups experienced the prototype (the coach, the exercises, or the interaction methods) in different ways. This is an important aspect of designing for these demographics. Addressing both, we intend larger and more varied participant samples in future studies.

8.3 Conclusions

Previous caveats notwithstanding, our results do indicate that our digital coach has significant beneficial effects for low-literate learners (using VESSEL). We mention in Sect. 1 that an ECA coach could benefit low literates by ‘putting a human face on computer support’. The results from this study show that the low-literate users accepted our coach as a useful source of help that could be relied on. Real interaction was observed between participants and coach: Help was asked for, offered, and accepted, and a small number of participants actually engaged the coach in dialogue, suggesting that as predicted by Bickmore & Picard [15], a friendly relationship of trust has started to form. Miao et al. [67] have already shown that ECA coaches in general can be accepted by learners; our work extends on this by showing that our ECA coach design is accepted by low-literate learners in particular. These positive effects were particularly seen with participants who struggled with the exercises, suggesting that they were helped the most by the coach’s presence and support. Since our primary goal with VESSEL is to support exactly these learners (learners in Kurvers et al.’s [56] profiles 2 and 3, see Sect. 6.3), this is promising. All the same, we do note that these positive effects were only found in the *Hard online banking* exercise, which was designed to test information skills. We clearly show that the coach supports information skills learning, but do not (yet) show a similar benefit to communication skills learning.

In conclusion, the starting assumption of our work (that a carefully designed virtual coach with integrated cognitive/affective/social learning support would work with low-literate societal participation learners) is confirmed, opening the possibility for more specialized work in this area. Our own future work will build on these results. The following iteration in our VESSEL design process will focus on addressing the issue of unstructured rules described above, taking the other study pitfalls and learned lessons into account. Now that the proof-of-concept evaluation has shown the validity of the core VESSEL ideas, we intend to create a formally structured VESSEL design specification, that comprehensively describes how to create situated interactive exercises at the right level of difficulty, and how to structure learning support such that an ECA coach can accurately provide it without requiring a human operator.

A Appendix: requirements derivation

This appendix presents a description of all new requirements shown in Table 1, and literature-backed rationales explaining why each generic requirement was refined for the coach, for the exercises, or for both.

Requirement *R1. Adaptability* is refined for both coach and exercises. The coach should tailor its interaction with the user to that individual user’s needs, wishes, and learning goals (*R1.1-C*). The perceived level of difficulty of any exercise could be altered by giving support quicker and in more detail. This (perceived) difficulty level should not be too easy and not be too hard, but instead fall in the Zone of Proximal Development ([106]; cf. [81, 103]). And over time, the coach can build a user model of individual learners and adapt its offered support even more closely. Individual exercises should not change their difficulty level mid-practice, as exercises are carefully designed with specific learning goals, tasks, and challenges. Rather, adaptability should be reached by building a corpus of different exercises that span a range of difficulty levels (*R1.1-E*). Note that these difficulty levels could be affected by other requirements. A particularly difficult exercise might not at all be sensitive (*R2*) or multimodal (*R5*) if the exercise’s learning goal is to teach learners to deal with crucial practical situations that are normally challenging for these reasons.

Requirement *R2. Sensitivity* is also refined for both coach and exercises. The coach’s dialogue and interaction style should be written from a sensitive point of view (*R2.1-C*). The coach should be calm and kind, and avoid saying things that upset low-literate users (for instance, being dismissive of their reading and writing problems). This lets the coach present exactly the ‘human face of support’ that low-literate users want (cf. [88]) By using sensitivity to convey empathy and build trust, the coach can encourage low-literate users to engage with it and accept offered learning support. Exercises should structure and portray their content as sensitively as needed to support desired learning goals and difficulty levels (*R2.1-E*).

Requirement *R3. Situatedness* is refined for exercises only. Exercises should be situated in the crucial practical situations and daily experiences of low-literate users (*R3.1-E*). Note that this demand for situated correctness can override the need for sensitivity (*R2*) and multimodality (*R5*) and in this way strongly determine exercise difficulty (*R1*). For instance, an exercise involving reading a difficult text (e.g. online banking) should not be multimodal.

Refinement of requirement *R5. Multimodality* applies to both coach and exercises. The coach should primarily use audio ‘speech’, and supplement this with visual or textual supporting material (*R5.1-C*). This fits both the ‘audio’ preference of low literates that primarily struggle with reading and writing, and the ‘slow reading’ preference of low literates that primarily struggle with rapid speech comprehension [56, 88]. Exercises should be as multimodal as needed to achieve their desired learning goals and difficulty level (*R5.1-E*).

Refinement of requirement *R6. Support* applies to exercise- and learner-specific learning support offered by the

coach, which takes three shapes: Cognitive support (R6.1-C), affective support (R6.2-C), and social support (R6.3-C). The coach should provide this learning support as described in Sect. 3.2.

Requirement R7. *Interactivity* again applies to both coach and exercises. The coach should interact with learners actively and passively. Actively, the coach should monitor the user's exercise progress (R7.1-C). If problems are detected, the coach should offer help. Passively, the coach should be able to reply to user questions and comments (R7.2-C). User comments should be acknowledged, and user questions should be answered to the best of the coach's ability. The exercises should be practical, interactive skills-training exercises: Learners should be required to use system input mechanics (e.g. mouse, keyboard, speech) to actively complete them (R7.1-E).

Finally, requirement R8. *Gaming principles* applies to the coach only, who should use gaming principles to invoke pride and a sense of achievement in the user (R8.1-C). This does not necessarily mean using actual game elements. The coach should focus on emphasizing and highlighting the user's successes and accomplishments, even when discussing problems.

B Appendix: use cases

Two use cases are provided here, based on the online banking and service desk exercises (Sect. 3.3). These use cases describe the envisioned optimal way that a typical low-literate learner would interact with VESSEL and make explicit how the requirements are met by the prototype. Use cases should be read as follows. *Pre-conditions* list the conditions that are necessarily true at the start of the use case. *Action sequence* describes the actions taken by the actors over the course of the use case. Particular action sequence steps may reference the requirements described in Table 1, indicating that this step demonstrates that requirement. Finally, *post-conditions* lists the measurable desired outcomes that we claim result from the action sequence. These claims are derived from the nine system objectives in Sect. 1. Only claims associated with (cognitive, affective, and social) learning experience and learning outcomes are used here; since users are pre-assumed to be working with VESSEL, accessibility claims cannot be tested. In the use cases, 'user' refers to the low-literate learner using the VESSEL system for learning purposes. 'Coach' refers to the ECA that provides cognitive, affective, and social learning support. 'Conversation partner' refers to the 'city hall service desk employee' character used in the service desk scenario.

B.0.1 UC1: online banking exercise

This use case describes an example online banking exercise that focuses on reading, writing, and information use skills. The goal of the exercise is to use an online banking website to transfer money from one account to another. The coach supports the user with cognitive, affective, and social learning support.

Pre-conditions

1. The user is interacting with the coach-supported VESSEL system.
2. An online banking exercise has been selected (either *Easy online banking* or *Hard online banking*).
3. The digital coach and the online banking website are both visible to the user.

Action sequence

1. The coach briefly introduces the goal and the scope of the exercise to the user. Example sentence: "*In this exercise, you will use online banking to transfer money to a webshop.*" (R3.1-E)
2. The coach makes small talk with the user about the topic of the exercise, online banking. It asks the user questions about their experiences with online banking, and tells the user about its own 'experiences'. Example sentences: "*Have you ever used online banking before?*" "*I used to think online banking was intimidating. What do you think?*" (R6.3-C)
3. The coach signals to the user that they can now start doing the exercise. The user starts doing the exercise. (R7.1-E)
4. The user attempts to navigate to the correct page on the online banking website, but does not know which page is the right one. The coach uses verbal scaffolding techniques to provide the user with the needed level of help. Example sentences: "*Do you know what you are supposed to be doing?*" "*You need to find the page where you can transfer money. Do you know where this page is?*" "*Click on the word 'online banking' to proceed in the exercise.*" (R1.1-C, R5.1-C, R6.1-C, R7.1-C)
5. The user attempts to fill out needed information on the correct page, but does not know what information to fill out where. The user asks the coach for input. The coach uses verbal scaffolding techniques to provide the user with the needed level of help. Example sentences: "*Do you know where to fill out your bank account number?*" "*You fill out your bank account number in the box labeled 'IBAN'. Do you know what 'IBAN' means?*" (R1.1-C, R5.1-C, R6.1-C, R7.2-C)

6. The user successfully completes the exercise. The coach signals to the user that the exercise is completed. Example sentence: “*Well done! You have completed this exercise.*” (R1.1-E, R8.1-C)
7. The coach uses motivational interviewing techniques to help the user reflect on the exercise. The coach offers specific performance feedback, based on the user’s speed and accuracy. Example sentence: “*I see you had trouble completing this exercise. But you took the time needed to complete it correctly.*” The coach also asks the user’s opinion on the design of the online banking website, and offers its ‘own’ complementing opinion. Example sentences: “*What did you think of this website?*” “*I thought this website was very confusing.*” (R2.1-C, R6.2-C)
8. The coach signals to the user that this particular training session is now over. Example sentence: “*We are done with practicing for now.*”

Post-conditions

1. The user performed well and has carried out the online banking steps correctly.
2. The user had a positive experience while doing the online banking exercise.
3. The user engaged with the coach on the topic of online banking.
4. The user has successfully completed the online banking exercise.
5. The user’s self-efficacy with regard to online banking has been increased.
6. The user is more motivated to learn about online banking in their daily life.

B.0.2 UC2: service desk exercise

This use case describes an example service desk exercise that focuses on speaking, understanding, and general communication skills. The goal of this exercise is to speak to a city hall employee at a service desk, to report the loss of the user’s passport and request a replacement. The conversation partner plays the role of the city hall employee. The coach supports the user with cognitive, affective, and social learning support.

Pre-conditions

1. The user is interacting with the coach-supported VESSEL system.
2. A service desk exercise has been selected (either *Easy service desk* or *Hard service desk*).
3. The digital coach and the conversation partner are both visible to the user.

Action sequence

1. The coach briefly introduces the goal and the scope of the exercise to the user. Example sentence: “*In this exercise, you will talk to a service desk employee and tell them about your lost passport.*” (R3.1-E)
2. The coach makes small talk with the user about the topic of the exercise, formal conversations (at city hall). It asks the user questions about their experiences with having formal conversations, and tells the user about its own ‘experiences’. Example sentences: “*Have you ever had to talk to someone at city hall before?*” “*People at service desks often use complicated language. I think that makes it difficult to understand.*” (R6.3-C)
3. The coach signals to the user that they can now start doing the exercise. The user starts doing the exercise. (R7.1-E)
4. The conversation partner starts the exercise by asking the user questions. When the user answers these questions correctly, the conversation partner moves the exercise along. Example sentences: “*Hello! Welcome to city hall. What can I help you with?*” “*Oh, I am sorry to hear you lost your passport. I will help you report this. What is your first name?*” (R2.1-E, R3.1-E, R5.1-E, R7.1-E)
5. The user attempts to answer the conversation partner’s questions, but they do not always understand what is being asked. The coach uses verbal scaffolding techniques to provide the user with the needed level of help. Example sentences: “*Do you know what this person is asking of you?*” “*This person is asking you what your social security number is. Do you know what this means?*” “*You can find your social security number on your passport, ID card, or driver’s license.*” (R1.1-C, R5.1-C, R6.1-C, R7.1-C)
6. The user successfully completes the exercise. The coach signals to the user that the exercise is completed. Example sentence: “*Well done! You have completed this exercise.*” (R1.1-E, R8.1-C)
7. The coach uses motivational interviewing techniques to help the user reflect on the exercise. The coach offers specific performance feedback, based on the user’s speed and accuracy. Example sentence: “*You completed the exercise really quickly! I see you still made some errors. Take your time to do the exercise flawlessly next time.*” The coach also asks the user’s opinion on the attitude and helpfulness of the ECA conversation partner, and offers its ‘own’ opinion. Example sentences: “*What did you think of the woman behind the service desk?*” “*I thought she was very friendly and did a good job trying to help you.*” (R2.1-C, R6.2-C)
8. The coach signals to the user that this particular training session is now over. Example sentence: “*We are done with practicing for now.*”

Post-conditions

1. The user performed well and held the service desk conversation correctly.
2. The user had a positive experience while doing the service desk exercise.
3. The user engaged with the coach on the topic of service desk situations.
4. The user has successfully completed the service desk exercise.
5. The user's self-efficacy with regard to formal service desk conversations has been increased.
6. The user is more motivated to learn about similar formal conversations in their daily life.

Acknowledgements The authors would like to thank Knup Fuhri and Mirella Verspik for their valuable contributions.

Declarations

Conflict of interest The work in this publication was supported by the 'Interaction For Universal Access' track of the Dutch national program COMMIT. On behalf of all authors, the corresponding author states that there is no conflict of interest.

Research involving Human Participants and/or Animals The work in this publication includes research with human participants. The research approach used in the work was evaluated and approved by the TU Delft Human Research Ethics Committee.

Informed consent Informed consent was obtained for all work described in this publication involving human participants. The informed consent form used in this work was evaluated and approved by the TU Delft Human Research Ethics Committee.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Alexander, I.F., Maiden, N.: Scenarios, Stories, Use Cases: Through the Systems Development Life-Cycle. John Wiley and Sons, New Jersey (2005)
2. Rekenkamer, A.: Aanpak van laaggeletterdheid. Tech. rep, Algemene Rekenkamer (2016)
3. Alsharbi, B., Richards, D.: Using virtual reality technology to improve reality for young people with chronic health conditions. In Proceedings of the 9th international conference on computer and automation engineering, pp. 11–15 (2017)
4. de Angeli, A., Brahmam, S.: Sex Stereotypes and Conversational Agents. AVI 2006 Gender and Interaction: Real and Virtual Women in a Male World Workshop paper (2006). www.brahnam.info/papers/EN2033.pdf
5. Bandura, A.: Self-Efficacy: The Exercise of Control. W.H. Freeman, New York (1997)
6. Barak, M.: Instructional principles for fostering learning with ICT: teachers perspectives as learners and instructors. Educ. Inf. Technol. **11**(2), 121–135 (2006). <https://doi.org/10.1007/s11134-006-7362-9>
7. Bass, K.M., Glaser, R.: Developing Assessments to Inform Teaching and Learning. Tech. rep. (2004). <https://eric.ed.gov/?id=ED483395>
8. Bayles, M.: Online banking: why people are branching out. Transfer **35**, 90 (2004)
9. Bercht, M., Viccardi, R.M.: Pedagogical agents with affective and cognitive dimensions. In: Proceedings of 5th Ibero-American Congress on Educational Informatics (RIBIE 2000). Nina del Mar, Chile (2000)
10. Berger, P.L., Luckmann, T.: The Social Construction of Reality: A Treatise in the Sociology of Knowledge (1966)
11. Bersee, T.: Veertig jaar aanpak laaggeletterdheid: ambitie, onvermogen en lange adem (2019)
12. Bickmore, T.W., Cassell, J.: In Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 396–402
13. Bickmore, T.W., Cassell, J.: Social dialogue with embodied conversational agents. Adv. Nat. Multim. Dial. Syst. **30**, 23–54 (2005)
14. Bickmore, T.W., Pfeifer, L., W. Jack, B.: Taking the Time to Care: Empowering Low Health Literacy Hospital Patients with Virtual Nurse Agents. In: CHI '09 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1265–1274 (2009)
15. Bickmore, T.W., Picard, R.W.: Establishing and maintaining long-term human-computer relationships. ACM Trans. Computer-Hum. Interact. **12**(2), 293–327 (2005)
16. Bickmore, T.W., Schulman, D., Sidner, C.L.: Automated interventions for multiple health behaviors using conversational agents. Patient Educ. Couns. **92**(2), 142–8 (2013). <https://doi.org/10.1016/j.pec.2013.05.011>
17. Henkemanns, O.A.B., van der Boog, P.J.M., Lindenberg, J., van der Mast, C.A.P.G., Neerincx, M.A., Zwetsloot-Schonk, B.J.H.M.: An online lifestyle diary with a persuasive computer assistant providing feedback on self-management. Technol. Health Care **17**(3), 253–267 (2009)
18. Bloom, B.S.: Taxonomy of Educational Objectives. The Cognitive Domain. Longman, Handbook I (1956)
19. Brinkman, W.P., van der Mast, C., Sandino, G., Gunawan, L.T., Emmelkamp, P.M.G.: The therapist user interface of a virtual reality exposure therapy system in the treatment of fear of flying. Interact. Comput. **22**(4), 299–310 (2010). <https://doi.org/10.1016/j.intcom.2010.03.005>
20. Buisman, M., Houtkoop, W.: Laaggeletterdheid in kaart. Tech. rep., Expertisecentrum Beroepsonderwijs & Stichting Lezen en Schrijven (2014)
21. Cassell, J., Bickmore, T.W.: External manifestations of trustworthiness in the interface. Commun. ACM **43**(12), 50–56 (2000)
22. Cassell, J., Bickmore, T.W.: Negotiated collusion: modeling social language and its relationship effects in intelligent agents. User Model. User-Adap. Inter. **13**, 89–132 (2003)
23. Cassell, J., Sullivan, J., Churchill, E., Prevost, S.: Embodied Conversational Agents. MIT Press, Cambridge (2000)

24. Cybinski, P., Selvanathan, S.: Learning experience and learning effectiveness in undergraduate statistics: modeling performance in traditional and flexible learning environments. *Decis. Sci. J. Innov. Educ.* **3**(2), 251–271 (2005). <https://doi.org/10.1111/j.1540-4609.2005.00069.x>
25. De Jong, M., Theune, M., Hofs, D.: Politeness and alignment in dialogues with a virtual guide. In: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1, pp. 207–214 (2008). <http://portal.acm.org/citation.cfm?id=1402416>
26. Deneka, A.A.: A Coach ECA to Increase Societal Participation of Low literate and Non-Native Citizens in the Societal Participation Learning Support System. Master thesis, University of Twente (2014)
27. Echevarria, J., Vogt, M., Short, D.: Making Content Comprehensible for English Learners: The SIOP Model. Allyn and Bacon, Boston (2004)
28. Fogg, B.J.: Persuasive Technology: Using Computers to Change What We Think and Do. Morgan Kaufman, San Francisco (2002)
29. Fortune, T.: Scaffolding Techniques in CBI Classrooms. Tech. rep. (2004)
30. Friederichs, S.A.H., Oenema, A., Bolman, C., Guyaux, J., van Keulen, H.M., Lechner, L.: I Move: systematic development of a web-based computer tailored physical activity intervention, based on motivational interviewing and self-determination theory. *BMC Publ. Health* **14**(1), 1–15 (2014)
31. Garrison, D.R., Archer, W.: A transactional perspective on teaching and learning: a framework for adult and higher education. Pergamon, Oxford (2000)
32. Gaver, W.W., Boucher, A., Pennington, S., Walker, B.: Cultural probes and the value of uncertainty. *Interactions-Funol.* **11**(5), 53–56 (2004)
33. Gaver, W.W., Dunne, T., Pacenti, E.: Cultural Probes. *Interactions* **6**(February), 21–29 (1999). <https://doi.org/10.1145/291224.291235>
34. Gijsberts, M., Hujink, W., Dagevos, J.: Jaarrapport integratie 2011. Tech. rep, Sociaal en Cultureel Planbureau, Den Haag (2012)
35. Glaser, B.G., Strauss, A.: The Discovery of Grounded Theory: Strategies for Qualitative Research. Transaction Publishers (1967)
36. de Greef, M.: Leren voor leven: een eigen plek in het dagelijks leven. Tech. rep, Spectrum GMO Gelderland, Velp (2009)
37. de Greef, M., Segers, M., Nijhuis, J.: Feiten & Cijfers geleterdheid. Tech. rep., Stichting Lezen & Schrijven (2014)
38. de Greef, M., Segers, M., Nijhuis, J.: Feiten & Cijfers Laageletterdheid: De invloed van lage basisvaardigheden op deelname aan de maatschappij. Maastricht University, Maastricht, Tech. rep. (2018)
39. de Greef, M., Segers, M., Nijhuis, J., Lam, J.F.: Impactonderzoek taaltrajecten Taal voor het Leven door Stichting Lezen & Schrijven op het gebied van sociale inclusie en leesvaardigheid. Maastricht University, Maastricht, Tech. rep. (2014)
40. Groot, A., Coppens, K., Lam, J.F.: Motiveren van laaggeletterden: Een literatuurstudie naar succesvolle interventies. ECBO (2019)
41. Gubbels, J., van Langen, A.M.L., Maassen, N.A.M., Meelissen, M.R.M.: Resultaten PISA-2018 in vogelvlucht. Tech. rep., Universiteit Twente, Enschede (2019). <https://doi.org/10.3990/1.9789036548226>
42. Haake, M., Gulz, A.: Visual realism and virtual pedagogical agents. In: Knight, J. (ed.) Proceedings of the 3rd International Design and Engagability Conference at NordiChi 2006 (iDec3) (2006). https://doi.org/10.1007/11821830_44
43. Hoeymans, N., Timmermans, J.M., de Klerk, M.M.Y., de Boer, A.H., Deeg, D.J.H., Poppelaars, J.L., Thissen, F., Fortuijn, J.C.D., de Hollander, A.E.M.: Gezond actief: de relatie tussen ziekten, beperkingen en maatschappelijke participatie onder Nederlandse ouderen. Tech. rep, RIVM (2005)
44. Hudlicka, E.: Virtual coach for mindfulness meditation training. In: 2011 AAAI Spring Symposium Series (2011)
45. Jonassen, D.H.: Objectivism versus constructivism: do we need a new philosophical paradigm? *Educ. Tech. Res. Dev.* **39**(3), 5–14 (1991)
46. Jonassen, D.H.: Thinking technology: toward a constructivist design model. *Educ. Technol.* **34**(4), 34–37 (1994)
47. Jonassen, D.H., Davidson, M., Collins, M., Campbell, J., Haag, B.B.: Constructivism and computer? Mediated communication in distance education. *Am. J. Distance Educ.* **9**(2), 7–26 (1995). <https://doi.org/10.1080/08923649509526885>
48. Kang, S.H., Gratch, J.: People like virtual counselors that highly disclose about themselves. *Ann. Rev. Cyberther. Telemed.* **16**7, 143–148 (2011)
49. Khan, R., De Angeli, A.: The attractiveness stereotype in the evaluation of embodied conversational agents. In: IFIP Conference on Human-Computer Interaction, pp. 85–97. Springer, Berlin (2009)
50. Kim, Y., Baylor, A.L.: Pedagogical agents as social models to influence learner attitudes. *Educ. Technol.* **47**(01), 23–28 (2007)
51. Knowles, M.S.: Andragogy, not pedagogy. *Adult Leadersh.* **16**(10), 350–352 (1968)
52. Knowles, M.S.: What is andragogy? In: The Modern Practice of Adult Education: From Pedagogy to Andragogy. Revised and updated., chap. 4, pp. 40–59. Prentice Hall Regents, Englewood Cliffs, New Jersey (1970)
53. Knowles, M.S.: The Adult Learner: A Neglected Species. 1. Gulf Publishing Company, Houston, Texas (1973)
54. Kramer, L.L., ter Stal, S., Mulder, B.C., de Vet, E., van Velsen, L.: Developing embodied conversational agents for coaching people in a healthy lifestyle: Scoping review. *J. Med. Internet Res.* **22**(2), 1–11 (2020). <https://doi.org/10.2196/14058>
55. Kurvers, J., van de Craats, I.: Literacy and second language in the low countries. In: Young-Scholten, M. (ed.) Low-Educated Second Language and Literacy Acquisition Proceedings of the Third Annual Forum, pp. 17–23. Roundtuit Publishing, Newcastle upon Tyne, England, Durham (2007)
56. Kurvers, J., Dalderop, K., Stockmann, W.: Cursistprofielen Laageletterdheid NT1 & NT2. Tech. rep, Steunpunt Taal en Rekenen VE, Tilburg (2013)
57. Lane, H.C., Cahill, C., Foutz, S., Auerbach, D., Noren, D., Lussenhop, C., Swartout, W.: The effects of a pedagogical agent for informal science education on learner behaviors and self-efficacy. In: Yacef, K., Lane, H.C., Mostow, J., Pavlik, P. (eds.) Artificial Intelligence in Education: 16th International Conference, pp. 309–318 (2013)
58. Leavitt, M.O., Schneiderman, B.: Research-based web design & usability guidelines. Tech. rep. (2006)
59. Lester, J.C., Converse, S.A., Kahler, S.E., Barlow, S.T., Stone, B.A., Bhogal, R.S.: The persona effect: affective impact of animated pedagogical agents. In: Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (1997). <http://dl.acm.org/citation.cfm?id=258797>
60. Liaw, S.S.: Investigating students' perceived satisfaction, behavioral intention, and effectiveness of e-learning: A case study of the Blackboard system. *Comput. Educ.* **51**(2), 864–873 (2008). <https://doi.org/10.1016/j.compedu.2007.09.005>
61. Looije, R., Neerinx, M.A., Cnossen, F.: Persuasive robotic assistant for health self-management of older adults: design

- and evaluation of social behaviors. *Int. J. Hum Comput Stud.* **68**(6), 386–397 (2010)
62. ter Maat, M., Truong, K.P., Heylen, D.: How agents turn-taking strategies influence impressions and response behaviours. *Presence Teleoper. Virtual Environ.* **20**(5), 412–430 (2011)
 63. Mahmood, A.K., Ferneley, E.: The use of animated agents in e-learning environments: an exploratory, interpretive case study. *ALT-J Res. Learn. Technol.* **14**(2), 153–168 (2006). <https://doi.org/10.1080/09687760600668552>
 64. Maulsby, D., Greenberg, S., Mander, R.: Prototyping an intelligent agent through Wizard of Oz. In: *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp. 277–284. Amsterdam (1993). <https://doi.org/10.1145/169059.169215>
 65. Mezirow, J.: Transformative learning: theory to practice. In: *New Directions for Adult and Continuing Education* vol. 74, pp. 5–12 (1997)
 66. Mezirow, J.: Learning to Think Like an Adult: Core Concepts of Transformation Theory. In: Jossey-Bass, S.F. (ed.) *Learning as Transformation. Critical Perspectives on a Theory in Progress*, chap. 1, pp. 3–33 (2000)
 67. Miao, Y., Engler, J., Giemza, A., Weinbrenner, S., Hoppe, H.U.: Development of a process-oriented scaffolding agent in an open-ended inquiry learning environment. *Res. Pract. Technol. Enhanc. Learn.* **7**(2), 105–128 (2012)
 68. Miller, W.R., Rollnick, S.: Ten things that motivational interviewing is not. *Behav. Cogn. Psychother.* **37**(2), 129–140 (2009). <https://doi.org/10.1017/S1352465809005128>
 69. Mislevy, R.J., Oranje, A., Bauer, M.I., von Davier, A., Hao, J., Corrigan, S., Hoffman, E.: *Psychometric Considerations in Game-based Assessment*. GlassLab (2014)
 70. Moreno, R., Flowerday, T.: Students' choice of animated pedagogical agents in science learning: a test of the similarity-attraction hypothesis on gender and ethnicity. *Contemp. Educ. Psychol.* **31**(2), 186–207 (2006). <https://doi.org/10.1016/j.cedpsych.2005.05.002>
 71. Moreno, R., Mayer, R.E., Spires, H.A., Lester, J.C.: The case for social agency in computer-based teaching: do students learn more deeply when they interact with animated pedagogical agents. *Cogn. Instr.* **19**(2), 177–213 (2001). <https://doi.org/10.1207/S1532690XCI1902>
 72. Nass, C., Isbister, K., Lee, E.J.: Truth is beauty: Researching embodied conversational agents. *Embodied Conversat. Agents* (2000). <https://doi.org/10.1017/CBO9781107415324.004>
 73. Nass, C., Steuer, J., Tauber, E.R.: Computers are social actors. Conference companion on Human factors in computing systems - CHI '94 p. 204 (1994). <https://doi.org/10.1145/259963.260288>
 74. Neerincx, M., Vught, W., Blanson Henkemans, O., Oleari, E., Broekens, J., Peters, R., Kaptein, F., Demiris, Y., Kiefer, B., Fumagalli, D., et al.: Socio-cognitive engineering of a robotic partner for child's diabetes self-management. *Front. Robot. AI* **6**, 118 (2019)
 75. Neerincx, M.A.: Situated cognitive engineering for crew support in space. *Pers. Ubiquitous Comput.* **15**(5), 445–456 (2011). <https://doi.org/10.1007/s00779-010-0319-3>
 76. Neerincx, M.A., Lindenberg, J.: Situated cognitive engineering for complex task environments. In: Schraagen, J.M., Militello, L.G., Ormerod, T., Lipshitz, R. (eds.) *Naturalistic Decision Making and Macrocognition*, pp. 373–390. Ashgate Publishing Limited, Aldershot (2008)
 77. Nielsen, J.: Ten usability heuristics. *Commun. ACM* **3**(1990), 1–2 (2005). <https://doi.org/10.1145/259963.260531>
 78. OECD: Literacy in the Information Age: Final Report of the International Adult Literacy Survey. Tech. rep., Statistics Canada (2000)
 79. Pauwels, G., Pickery, J.: *Wie participeert niet? Ongelijke deelname aan het maatschappelijke leven in verschillende domeinen*. SVR-rapport 2007/5 (2007)
 80. Peeters, M.M.M.: *Personalized Educational Games - Developing Agent-supported Scenario-based Training*. Ph.D. thesis (2014)
 81. Peeters, M.M.M., van den Bosch, K., Meyer, J.J.C., Neerincx, M.A.: The design and effect of automated directions during scenario-based training. *Comput. Educ.* **70**, 173–183 (2014). <https://doi.org/10.1016/j.compedu.2013.07.039>
 82. Peeters, M.M.M., Meyer, J.J.C., van den Bosch, K., Neerincx, M.A.: Situated Cognitive Engineering: The Requirements and Design of Automatically Directed Scenario-based Training. In: *Proceedings of ACHI 2012: The Fifth International Conference on Advances in Computer-Human Interactions*, c, pp. 266–272 (2012)
 83. van de Pol, J., Elbers, E.: Scaffolding student learning: a micro-analysis of teacher-student interaction. *Learn. Cult. Soc. Interact.* **2**(1), 32–41 (2013). <https://doi.org/10.1016/j.lcsi.2012.12.001>
 84. Reeves, B., Nass, C.: *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, New York (1996)
 85. Richards, C.: The design of effective ICT-supported learning activities: exemplary models, changing requirements, and new possibilities. *Lang. Learn. Technol.* **9**(1), 60–79 (2005)
 86. de Rosis, F., Novielli, N., Carofiglio, V., Cavalluzzi, A., De Carolis, B.: User modeling and adaptation in health promotion dialogs with an animated character. *J. Biomed. Inf.* **39**(5), 514–31 (2006). <https://doi.org/10.1016/j.jbi.2006.01.001>
 87. Schouten, D.G.M.: *A virtual coach for low-literates to practice societal participation*. ProefschriftMaken (2020)
 88. Schouten, D.G.M., Paulissen, R.T., Hanekamp, M., Groot, A., Neerincx, M.A., Cremers, A.H.M.: Low-literates' support needs for societal participation learning: Empirical grounding of theory- and model-based design. *Cognit. Syst. Res.* **45**, 30–47 (2017). <https://doi.org/10.1016/j.cogsys.2017.04.007>
 89. Schouten, D.G.M., Smets, N.J.J.M., Driessen, M., Fuhri, K., Neerincx, M.A., Cremers, A.H.M.: Requirements for a virtual environment to support the social participation education of low-literates. *Univ. Access Inf. Soc.* **16**(3), 681–698 (2016). <https://doi.org/10.1007/s10209-016-0502-z>
 90. Schouten, D.G.M., Smets, N.J.J.M., Driessen, M., Hanekamp, M., Cremers, A.H.M., Neerincx, M.A.: User requirement analysis of social conventions learning applications for non-natives and low-literates. In: *Engineering Psychology and Cognitive Ergonomics. Understanding Human Cognition*, pp. 354–363. Springer (2013)
 91. Schouten, D.G.M., Venneker, F., Bosse, T., Neerincx, M.A., Cremers, A.H.M.: A digital coach that provides affective and social learning support to low-literate learners. *IEEE Trans. Learn. Technol.* **11**(1), 67–80 (2017). <https://doi.org/10.1109/TLT.2017.2698471>
 92. Shamekhi, A., Trinh, H., Bickmore, T.W., DeAngelis, T.R., Ellis, T., Houlihan, B.V., Latham, N.K.: A virtual self-care coach for individuals with spinal cord injury. In: *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 327–328. ACM (2016)
 93. Shaw, E., Ganesan, R., Johnson, W.L., Millar, D.: Building a case for agent-assisted learning as a catalyst for curriculum reform in medical education. *Front. Artif. Intell. Appl.* **50**, 509–516 (1999)
 94. Shea, P., Fredericksen, E., Pickett, A., Pelz, W., Swan, K.: Measures of learning effectiveness in the SUNY learning network. *Online Educ.* **2**, 31–54 (2001)
 95. Shute, V.J., Graf, E.A., Hansen, E.G.: Designing adaptive, diagnostic math assessments for individuals with and without visual disabilities. Tech. Rep. 01 (2006)

96. Shute, V.J., Hansen, E.G., Almond, R.G.: An Assessment for Learning System Called ACED: Designing for Learning Effectiveness and Accessibility. *ETS Res. Rep. Ser.* **2007**(2), i–45 (2007)
97. Shute, V.J., Zapata-Rivera, D.: Adaptive educational systems. *Adaptive technologies for training and education* **7**(27) (2012)
98. Skalski, P., Tamborini, R.: The role of social presence in interactive agent-based persuasion. *Media Psychol.* **10**(3), 385–413 (2007)
99. Sobell, L., Sobell, M.: Motivational interviewing strategies and techniques: rationales and examples (2008)
100. Raad, Sociaal-Economische.: Samen werken aan taal: Een advies over laaggeletterdheid. *Tech. Rep.* April, Sociaal-Economische Raad, Den Haag (2019)
101. ter Stal, S., Kramer, L.L., Tabak, M., den Akker, H., Hermens, H.: Design features of embodied conversational agents in eHealth: a Literature Review. *Int. J. Human-Comput. Stud.* **138**, 102409 (2020). <https://doi.org/10.1016/j.ijhcs.2020.102409>
102. ter Stal, S., Tabak, M., den Akker, H., Beinema, T., Hermens, H.: Who do you prefer? the effect of age, gender and role on users' first impressions of embodied conversational agents in ehealth. *Int. J. Human-Comput. Interact.* **36**(9), 881–892 (2020)
103. Vacca, J.S.: Using scaffolding techniques to teach a social studies lesson about buddha to sixth graders. *J. Adolesc. Adult Lit.* **51**(8), 652–658 (2008). <https://doi.org/10.1598/JAAL.51.8.4>
104. Veletsianos, G., Miller, C.: Conversing with pedagogical agents: a phenomenological exploration of interacting with digital entities. *Br. J. Edu. Technol.* **39**(6), 969–986 (2008)
105. van Vugt, H.C., Konijn, E.A., Hoorn, J.F., Veldhuis, J.: Why fat interface characters are better e-health advisors. In: *International Workshop on Intelligent Virtual Agents*, pp. 1–13. Springer, Berlin (2006)
106. Vygotsky, L.S.: *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge (1980)
107. Westera, M., Boschloo, J., van Diggelen, J., Koelewijn, L.S., Neerinx, M.A., Smets, N.J.J.M.: Employing use-cases for piecewise evaluation of requirements and claims. In: *Proceedings of the 28th Annual European Conference on Cognitive Ergonomics*, p. 8 (2010). <http://dl.acm.org/citation.cfm?id=1962357>
108. Zhang, D., Zhou, L., Briggs, R.O., Nunamaker, J.F.: Instructional video in e-learning: assessing the impact of interactive video on learning effectiveness. *Inf. Manage.* **43**(1), 15–27 (2006). <https://doi.org/10.1016/j.im.2005.01.004>
109. Zhou, S., Zhang, Z., Bickmore, T.: Adapting a persuasive conversational agent for the chinese culture. In: *2017 International Conference on Culture and Computing (Culture and Computing)*, pp. 89–96. IEEE (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.