A virtual coach for low-literates to practice societal participation

Schouten, D.G.M.

DOI
10.4233/uuid:7f98f3d8-bebc-4927-8b5d-d70c84bfa04c

Publication date
2020

Document Version
Final published version

Citation (APA)
Schouten, D. G. M. (2020). *A virtual coach for low-literates to practice societal participation*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:7f98f3d8-bebc-4927-8b5d-d70c84bfa04c

A virtual coach for low-literates to practice societal participation

A virtual coach for low-literates to practice societal participation

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus, prof. dr. ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on ***Thursday 17 December 2020*** at ***15:00 uur*** o'clock

by

Dylan Gerard Michel SCHOUTEN

Master of Science in Human-Technology Interaction
Eindhoven University of Technology, Eindhoven, the Netherlands
Born in Tilburg, the Netherlands

This dissertation has been approved by the promotors.

promotor: prof. dr. M.A. Neerincx
copromotor: dr. A.H.M. Cremers

Composition of the doctoral committee:

Rector Magnificus,	Chairperson
Prof. dr. M.A. Neerincx,	Delft University of Technology, TNO
Dr. A.H.M. Cremers,	HU University of Applied Sciences Utrecht, TNO
Independent members:	
Prof. dr. D.K.J. Heylen,	Universiteit Twente
Prof. dr. M. de Greef,	Vrije Universiteit Brussel, België
Prof. dr. W.A. IJsselsteijn,	Technische Universiteit Eindhoven
Prof. dr. M.M. Specht,	Technische Universiteit Delft
Prof. dr. F.M. Brazier,	Technische Universiteit Delft

Keywords: Societal participation; Low-literacy; Virtual learning environment; Socio-Cognitive Engineering; Requirements engineering; Qualitative methods

Printed by: ***ProefschriftMaken*** (proefschriftmaken.nl)

Front & Back: Illustration: Manu Adavi. Cover design and layout: Bregje Jaspers (proefschriftontwerp.nl).

ISBN 978-94-6423-079-6

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

Here’s to everyone who told me I could do it.
Turns out you were right after all.

Contents

Summary

This thesis presents the research, design, and evaluation of the learning support system VESSEL: *Virtual Environment to Support the Societal participation Education of Low-literates*. The project was started from the premise that people of low literacy in the Netherlands participate in society less often and less effectively than literate people do: Their lower ability to read, write, speak, and understand the Dutch language hampers their ability to independently be part of society. Our goal was to create learning support prototypes with a re-usable design rationale, aimed at helping these people of low literacy learn to improve their societal participation. To achieve this, low-literate learners participated throughout the entire design process, ensuring that we addressed their wants and needs with regard to learning and the perceived shortcomings of existing learning materials and kept in mind their skills and capabilities in order to ensure effective learning. Particularly, we investigated the possible ways that digital learning, Virtual Learning Environments (VLE), and Embodied Conversational Agents (ECA) could help fulfill the societal participation needs of this target group. We used the Socio-Cognitive Engineering (SCE) methodology to organize and structure this research, distinguishing the *foundation*, *specification* and *evaluation* of the VESSEL design. Two studies provided a grounded foundation for VESSEL, which was refined and worked out into three subsequent studies that provided the consequential design specifications and prototype evaluations (all prototypes have been tested with a human 'Wizard of Oz' simulating VESSEL functionality).

In the first study, we collected necessary information for the *foundation* of VESSEL in three categories. The first category consisted of the operational demands, which form an overview of the context of use: Demographic information about low-literate learners in the Netherlands, a description of the crucial practical situations of participating in Dutch society, and important attributes of learning societal participation in the Netherlands. The second category encompassed human factors knowledge, consisting of literature about adult learning and ICT-supported learning. The third category contained technology insights, which we gathered by looking at both existing learning support software in the areas of low literacy and societal participation in the Netherlands, and the envisioned capabilities of VLEs.

In the second study, we extended and refined our knowledge of the operational demands (as the *foundation* for VESSEL). We spoke to low-literate language learners in the Netherlands, in order to gain qualitative insights into their daily life experiences related to participating in Dutch society. We used participant workshops and Cultural Probes to obtain large amounts of rich data pertaining to these experiences, and we used the Grounded Theory method to transform these data into the *Societal Participation Experiences of Low-Literates* (SPELL) model. This model describes the four attribute categories of societal participation experiences:

Personal attributes, formal societal attributes, information societal attributes, and information-communication attributes.

In the third study, we used our foundation of information to create a *first prototype*, a 'proof-of-concept' VESSEL. This prototype consisted of four interactive scenario-based learning exercises: Two exercises ('Easy' and 'Hard') about conducting online banking, and two exercises about talking to a city hall service desk employee. The prototype also contained our ECA 'coach', Anna, who could provide three types of learning support: Cognitive learning support based on scaffolding, affective learning support based on motivational interviewing, and social learning support based on small talk. This prototype was evaluated with low-literate language learners throughout the Netherlands in an empirical mixed-method experiment, in which users did all four exercises both with and without coach support. Results showed that all learners managed to complete all exercises with coach support, while almost no learners completed all exercises without coach support. Participants interacted with the coach in a natural manner: They asked for her help and even talked to her without external prompting. A majority of participants appreciated her presence and help.

In the fourth study, we formalized the coach's cognitive learning support capabilities for the design and evaluation of the *second prototype*. We used existing scaffolding literature and our own experiences from the third study to define five levels of cognitive learning support: Prompt, Explanation, Hint, Instruction, and Modeling. We created a large corpus of standardized speech utterances for the coach in the context of the Hard Online Banking exercise, and wrote detailed rules describing which type of utterance the coach should use in any given situation, how long the coach should wait between utterances, and what kinds of user-uttered keywords she should react to and how. The model describes that the coach should always offer the lowest level of support (Prompt) for any new topic, that support should always go up in level and never repeat itself unless asked, and that the coach should wait a certain amount of time after any utterance. Two support models were made to describe this timing: The Generalized Model, in which the coach always waits 20 seconds, and the Individualized Model, in which the coach adapts the support wait time to the individual participant's previous performance. The second prototype was created, focusing on an expanded version of the Hard Online Banking exercise, and an empirical mixed-method experiment was conducted with low-literate learners to test the differences between the two support models: Learners completed three exercises in either the Generalized condition (with a consistent 20 second support wait time) or the Individualized condition (in which their support time in exercises two and three depended on their results in exercises one and two). We hypothesized that both the Generalized and Individualized Models would increase learning effectiveness, and that the Individualized Model would increase learning effectiveness significantly more than the Generalized Model. Results support the first hypothesis: Support from either model resulted in high learning effectiveness and higher self-efficacy for low-literate learners, and low-literate learners managed to use the new keyword-based speech recognition without the need for explanation. The second hypothesis was not supported: No differences in learning effectiveness were found between the two support models.

In the fifth study, we formalized the coach's affective and social support capabilities for the design and evaluation of the *third prototype*. We used existing motivational interviewing literature to define four levels of affective learning support (Reflective Listening, Normalizing, Affirmation, and Self-Efficacy Supporting) for three emotional states (Anger, Fear, and Sadness) at three levels of specificity (General, Specific, and Very Specific), and created a corpus of affective support utterances: For each combination of emotional state and specificity (General, Specific, or Very Specific Anger, Fear, or Sadness), one or two support utterances were created for each level of affective support. We used the *Shimmer* photoplethysmographic sensor and the *FaceReader* facial expression recognition software to infer learners' affective states from their heart rate and facial expressions (respectively), and connected this to new affective support rules: Whenever the coach inferred an emotion at a certain level of specificity, it should use one Reflective Listening utterance relevant to that particular combination, one Normalizing utterance, one Affirmation utterance, and one Self-Efficacy utterance, in that order. We also used existing small talk literature to write a simple branching small talk script for the coach, focused on bonding with the user and introducing the new Volunteer Work exercise, in which learners had to fill out a volunteer work background information form and then talk to an ECA about their answers. A third and final Wizard-of-Oz prototype was created and evaluated with low-literate learners in an empirical mixed-method experiment, in which learners completed the full exercise once with only cognitive learning support and once with cognitive, affective, and social learning support. Results did not show strong significant differences between the two conditions. We identified three potential explanations: Our exercises did not manage to evoke emotional reactions in learners strong enough for our sensors to detect, our affective support model was not effective in the way we intended, and/or our experimental setup limited the amount of emotional reactions learners could experience. However, the prototype in general did work as intended: Learners completed every exercise, requested and used the coach's support, and reported higher self-efficacy at the end of the experiment. This experiment also reported differences between NT1 and NT2 learners and between men and women, suggesting more careful study into demographic differences will be required.

Overall, results from our studies show that VESSEL seems to be increasing learning effectiveness. Learners across studies reported that working with VESSEL made for a positive learning experience, and after doing challenging societal participation exercises for the first time, learners' self-efficacy regarding the exercise topic (online banking / volunteer work) increased and stayed on the new level throughout. However, it proved difficult to clearly identify distinctive effects for specific VESSEL functions: For instance, positive learning outcomes could not clearly be attributed to the adaptive timing of the support or the constructive scaffolding used for cognitive support, and the positive experience of interacting with the coach could not be attributed to the presence of scripted small talk and affective support.

Crucially, our results show that learners were able to use VESSEL as intended: They interacted with the exercises as intended and with the coach as envisioned,

without the need for prior explanations or tutorials (save a brief introduction given by the coach). This suggests that we have managed to incorporate the actual capabilities, shortcomings, and wants and needs of people of low literacy into the design of VESSEL. However, it is not clear whether these positive outcomes would apply to all low-literate learners: While we attempted to recruit low-literate participants from different backgrounds and skill levels, on reflection, the majority of our participants were relatively high-skilled and intrinsically motivated. This is further complicated by the relatively low number of participants in our experiments, which calls the power of the results into question. Just as importantly, we regularly saw that learners socially engaged with 'Anna': They responded to her questions, asked questions of their own, thanked her for her help, and even occasionally talked to her as if she was a real person – telling stories and making jokes. Learners were grateful for the support, and generally indicated that they would like to receive more support like this in the future.

Samenvatting

Dit proefschrift beschrijft het onderzoeken, ontwikkelen, en evalueren van het leerondersteunings-programma VESSEL: *Virtual Environment to Support the Societal participation Education of Low-literates*. Dit project begon vanuit de stelling dat laaggeletterde mensen in Nederland minder vaak en minder effectief in de samenleving deelnemen dan geletterde mensen: Hun beperkte vermogen om Nederlands te lezen, schrijven, spreken, en begrijpen vermindert hun vermogen om zelfstandig mee te doen in de maatschappij. Ons doel was het ontwikkelen van leerondersteunings-prototypes met een ontwerp-rationele die herbruikbaar is, gericht op het helpen van laaggeletterde mensen die beter willen leren deelnemen. Om dit te bereiken zijn laaggeletterde deelnemers betrokken bij het gehele ontwerpproces, om ervoor te zorgen dat we hun leerbenodigdheden en -behoeften en de (ervaren) tekortkomingen van bestaand lesmateriaal benoemden en hun vermogens en vaardigheden in acht namen. We onderzochten in het bijzonder hoe digitaal leren, virtuele leeromgevingen ('*Virtual Learning Environments*', VLE), en '*Embodied Conversational Agents*' (digitale karakters met een menselijk uiterlijk, ECA) de maatschappelijke participatie-behoeften van deze doelgroep zouden kunnen helpen vervullen. We maakten gebruik van de *Socio-Cognitive Engineering*-methode (SCE) om het onderzoek vorm te geven en te organiseren, waarin onderscheid werd gemaakt tussen de '*foundation*', '*specification*' en '*evaluation*' van het ontwerp van VESSEL. Twee studies leverden een gegronde 'foundation' op voor VESSEL, die verder werd uitgewerkt en verfijnd in drie vervolgstudies die de belangrijke ontwerpspecificaties samenstelden en prototype-evaluaties evalueerden (alle prototypes zijn getest met een menselijke 'Wizard of Oz' die de functionaliteit van VESSEL simuleerde).

In de eerste studie hebben we noodzakelijke informatie voor de '*foundation*' van VESSEL verzameld in drie categorieën. De eerste categorie bestond uit de operationele eisen van de software, die de gebruikscontext omschrijven: Demografische informatie over laaggeletterde leerlingen in Nederland, een beschrijving van de cruciale praktijksituaties van Nederlandse participatie, en belangrijke attributen die deelname in de Nederlandse samenleving beschrijven. De tweede categorie omvatte '*human factors*'-kennis, bestaande uit literatuur over volwasseneneducatie en leren met ICT-ondersteuning. De derde categorie bevatte technologie-inzichten, die we hebben verzameld door te kijken naar zowel bestaande leerondersteunings-software voor laaggeletterde mensen, als de beoogde functionaliteiten van VLEs.

In de tweede studie hebben we onze kennis van de operationale eisen (als de '*foundation* ' van VESSEL) uitgebreid en verfijnd. We spraken met laaggeletterde Nederlandse taal-leerders, om kwalitatieve inzichten te krijgen in hun dagelijkse ervaringen met het deelnemen in de Nederlandse samenleving. We maakten gebruik van deelnemers-workshops en 'Cultural Probes' om een grote hoeveelheid rijke data te verzamelen over deze ervaringen, en van de Grounded Theory-methode om deze data te transformeren tot het *Societal Participation Experiences of Low-Literates* (SPELL)-model. Dit model beschrijft de vier attribuut-categorieën van participatie-ervaringen: Persoonlijke kenmerken, formele maatschappelijke kenmerken, informele maatschappelijke kenmerken, en informatie-communicatie-kenmerken.

In de derde studie hebben we onze 'foundation' van informatie gebruikt om een *eerste prototype* te maken, een 'proof-of-concept' van VESSEL. Dit prototype bestond uit vier interactieve scenario-gebaseerde leeroefeningen: Twee oefeningen ('Makkelijk' en 'Moeilijk') over internetbankieren, en twee oefeningen over praten met een baliemedewerker in het gemeentehuis. Het prototype bevatte ook onze ECA-'coach', Anna, die drie vormen van leerondersteuning aanbood: Cognitieve leerondersteuning gebaseerd op '*scaffolding*', affectieve leerondersteuning gebaseerd op motivationeel interviewen, en sociale leerondersteuning gebaseerd op '*small talk*'. Dit prototype is geëvalueerd door laaggeletterde taal-leerders door heel Nederland in een empirisch mixed-method experiment, waarin deelnemers allevier de oefeningen deden zowel met de leerondersteuning van de coach als zonder. De resultaten lieten zien dat alle leerlingen alle oefeningen af konden maken met de hulp van de coach, terwijl bijna geen enkele leerling zonder hulp van de coach alle oefeningen af kon ronden. Deelnemers communiceerden op een normale manier met de coach: Ze vroegen haar om hulp en praatten zelfs met haar zonder verdere aanmoediging. Een meerderheid van de deelnemers waardeerde haar hulp en aanwezigheid.

In de vierde studie hebben we de cognitieve leerondersteunings-vermogens van de coach geformaliseerd voor het ontwerpen en evalueren van het *tweede prototype*. We maakten gebruik van bestaande literatuur over scaffolding en onze eigen ervaringen uit de derde studie om vijf niveaus van cognitieve ondersteuning te definiëren: Prompt, Uitleg, Hint, Instructie, en Modelleren. We stelden een groot corpus van gestandaardiseerde spraak-uitingen samen voor de coach in de context van de Moeilijke Internetbankieren-oefening, en schreven gedetailleerde regels die uitleggen welke uiting de coach moet gebruiken in iedere mogelijke situatie, hoe lang de coach moet wachten tussen twee uitingen, en op welke sleutelwoorden in spraak-uitingen van gebruikers ze moet reageren, en hoe. Het model beschrijft dat de coach altijd het laagste niveau van ondersteuning (Prompt) aan moet bieden voor ieder nieuw onderwerp, dat ondersteuning altijd in niveau moet stijgen en zichzelf nooit moet herhalen tenzij hierom gevraagd wordt, en dat de coach een zekere hoeveelheid tijd moet wachten tussen uitingen. We hebben twee ondersteunings-modellen gemaakt om deze timing te beschrijven: Het Algemene Model, waarin de coach altijd 20 seconden wacht, en het Individuele Model, waarin de coach haar wachttijd aanpast aan de eerdere prestaties van de individuele leerling. Een tweede prototype, gefocust op een uitgebreidere versie van de Moeilijke Internetbankieren-oefening, is geëvalueerd in een empirisch mixed-method experiment met laaggeletterde taalleerders, om de verschillen tussen de twee modellen te testen: Leerders deden drie oefeningen in de Algemene conditie (met een consistente wachttijd van 20 seconden) of de Individuele conditie (waarin de wachttijd in de tweede en derde oefening afhing van hun prestatie in de eerste en tweede oefening). Onze hypotheses waren dat het Algemene Model en het Individuele Model allebei de leereffectiviteit zouden verhogen, en dat het Individuele Model de leereffectiviteit significant meer zou verhogen dan het Algemene Model. Onze resultaten ondersteunen de eerste hypothese: Ondersteuning van ieder model verhoogde de leereffectiviteit en de 'self-efficacy' van laaggeletterde leerlingen, en laaggeletterde leerlingen slaagden erin om de nieuwe sleutelwoord-gebaseerde spraakherkenning te gebruiken zonder directe uitleg. De tweede hypothese werd niet ondersteund: Er zijn geen verschillen gevonden in leereffectiviteit tussen de twee ondersteuningsmodellen.

In de vijfde studie hebben we de affectieve en sociale leerondersteuningsmogelijkheden van de coach geformaliseerd voor het ontwerpen en evalueren van het *derde prototype*. We maakten gebruik van bestaande literatuur over motivationeel interviewen om vier niveaus van affectieve ondersteuning te definiëren (Reflecterend Luisteren, Normaliseren, Bevestigen, en Self-Efficacy-Ondersteunen) voor drie emotionele staten (Woede, Angst, en Droefheid) in drie niveaus van nauwkeurigheid (Algemeen, Nauwkeurig, en Heel Nauwkeurig), en we hebben een corpus van affectieve ondersteuningsuitingen gemaakt: Voor elke combinatie van emotionele staat en nauwkeurigheid (Algemene, Nauwkeurige, of Heel Nauwkeurige Woede, Angst, of Droefheid) zijn een of twee ondersteuningsuitingen geschreven voor elk niveau van affectieve ondersteuning. We gebruikten de photoplethysmographische sensor *Shimmer* en het gezichtsuitdrukking-herkenningsalgoritme *FaceReader* om de affectieve staat van leerlingen af te kunnen leiden uit hun hartslag en gezichtsuitdrukking (respectievelijk), en verbonden dit aan nieuwe affectieve ondersteuningsregels: Op het moment dat de coach een emotie afleidde op een bepaald niveau van nauwkeurigheid, moet ze een Reflecteren Luisteren-uiting geven relevant voor deze specifieke combinatie, een Normaliseren-uiting, een Bevestigen-uiting, en een Self-Efficacy-Ondersteuning-uiting, in die volgorde. We maakten ook gebruik van bestaande literatuur over 'small talk' om een klein vertakkend script te schrijven voor de coach, dat zich richt op het scheppen van een band met de gebruiker en het introduceren van de nieuwe Vrijwilligerswerk-oefening, waarin leerlingen een formulier over vrijwilligerswerk moesten invullen en vervolgens met een ECA over hun keuzes moesten praten. Een derde en laatste Wizard-of-Oz prototype is geëvalueerd in een empirisch mixed-method experiment met laaggeletterde leerlingen, die de volledige Vrijwilligerswerk-oefeningen twee keer deden: Eén keer met alleen cognitieve ondersteuning, en één keer met cognitieve, affectieve, en sociale ondersteuning. De resultaten lieten geen significante verschillen zien tussen deze twee condities. We vonden drie mogelijke verklaringen: De emotionele reacties van leerlingen op onze oefeningen waren zo zwak dat onze sensoren ze niet konden herkennen, ons affectief ondersteuningsmodel werkte niet op de beoogde manier, en/of onze experimentele opzet beperkte de emotionele reacties die leerlingen konden ervaren. Het prototype werkte echter wel in het algemeen zoals bedoeld: Deelnemers konden iedere oefening afronden, vroegen om en maakten gebruik van de hulp van de coach, en toonden een hogere 'self-efficacy' aan het eind van het experiment. Dit experiment liet ook verschillen zien tussen NT1- en NT2-leerlingen en tussen mannen en vrouwen, wat suggereert dat nauwkeuriger onderzoek naar de invloed van demografische verschillen nodig is.

In het algemeen laten de resultaten van ons werk zien dat VESSEL de effectiviteit van leren verhoogt. In alledrie de studies rapporteren leerlingen dat werken met VESSEL de leerervaring positiever maakt, en na het voor de eerste keer doen van een moeilijke oefening stijgt de 'self-efficacy' van leerlingen met betrekking tot die oefening (internetbankieren / vrijwilligerswerk); deze hogere 'self-efficacy' blijft vervolgens consistent. Het is echter moeilijk gebleken om duidelijk verschillende effecten te identificeren voor specifieke VESSEL-functies: Positieve leeruitkomsten kunnen bijvoorbeeld niet overduidelijk worden toegeschreven aan de adaptieve timing van de ondersteuning of aan de opbouwende scaffolding die is gebruikt in de cognitieve ondersteuning, en de positieve ervaring van interacteren met de coach kan niet overduidelijk worden toegeschreven aan de aanwezigheid van een 'small talk'-draaiboek en affectieve ondersteuning.

Zeer belangrijk is dat onze deelnemers VESSEL konden gebruiken zoals wij hadden beoogd: Ze deden de oefeningen zoals beoogd en werkten samen met de coach zoals bedoeld, zonder lange uitleg of tutorials nodig te hebben (afgezien van een korte introductie gegeven door de coach). Dit suggereert dat het ons gelukt is om de daadwerkelijke vermogens, tekortkomingen, en wensen en behoeften van laaggeletterde mensen te verwerken in het ontwerp van VESSEL. Het is echter niet duidelijk of deze positieve uitkomsten waar zouden zijn voor alle laaggeletterden: Hoewel we geprobeerd hebben om laaggeletterde deelnemers te werven met verschillende achtergronden en vaardigheidsniveaus, kunnen we al terugkijkend zien dat de meerderheid van onze deelnemers relatief vaardig en intrinsiek gemotiveerd was. Dit wordt verder gecompliceerd door het relatief lage aantal deelnemers in onze experimenten, hetgeen de power van onze resultaten in twijfel trekt. Even belangrijk is dat we keer op keer zagen dat leerlingen sociaal interacteerden met 'Anna': Ze reageerden op haar vragen, stelden haar zelf vragen, bedankten haar voor haar hulp, en praatten zelfs af en toe gewoon met haar – ze vertelden verhalen en maakten grappen met haar. De leerlingen waren dankbaar voor de ondersteuning die door haar werd geboden, en gaven vaak aan dat ze in de toekomst meer van dit soort ondersteuning zouden willen hebben.

1

Introduction

”When I lost my job at the metalworks (…) I had to go to the Institute for Employee Insurance. And then that man says, here’s an envelope with everything you need for benefits… And you have to do that on the computer. So I say, on the computer? I can hardly read and write! (…) In my head, I saw my house of cards fall apart, bit by bit.”

Leo, literacy ambassador

1

1.1. Learning Support For Low-Literate Learners

Karel is a 45-year-old Dutch man, married with no children. He has completed lower vocational technical education, and works in a harbor depot in Rotterdam, while his wife stays at home. This week, however, Karel's wife is away on vacation with friends. And urgent bills have come in, which she usually takes care of. Since Karel works full-time and cannot visit his bank for help, the only way to pay these bills on time is to use online banking. But Karel has limited ability to read and write, and doesn't know how online banking works. He has opened the bank website once, but it was full of difficult words and many options and that scared him off. The only 'solution' Karel has is to wait for his wife to come back from vacation, so she can do it. That means the bills will be late, which will cause new problems, but what other options does he have?

Fatima is a 36-year-old Moroccan woman who migrated to the Netherlands 15 years ago to marry a Dutch man. Now she lives as a stay-at-home mother for their two children. Fatima is fluent in Moroccan, but not in Dutch, and she usually relies on her husband and children to translate for her. But now that her children are starting high school, she wants to become more independent. Fatima is especially interested in doing volunteer work, something she did in the past in Morocco to meet new people and help out around the neighbourhood. She knows the people in the neighbourhood center can help her get started with Dutch volunteer work. However, the Dutch language barrier is daunting. Her limited Dutch communication skills effectively limit her independence.

The stories of Karel and Fatima illustrate how challenging it is for people with limited reading, writing, speaking, and understanding skills to autonomously participate in society. These people are called *people of low literacy*, or *low-literate people*. *Societal participation* means acting in the context of a society in order to reach certain goals [1]. Societal participation encompasses a set of behaviours called *crucial practical situations* [2, 3]: This includes (e.g.) online banking and volunteer work, as well as legal matters, neighbourhood initiatives, going to school, and grocery shopping. Being able to participate autonomously means being able to handle these crucial practical situations successfully. In modern (digital) information societies, doing so requires mastery of information and communication skills: Reading, writing, speaking, and understanding that society's primary language [4–7]. Computer use and ICT skills have become a requirement as well [8]. The problems with societal participation that low-literate people encounter can be of a *cognitive* nature, such as Karel's inability to understand a website or Fatima's lack of experience with the Dutch language, but they can also be *affective*, like Karel's fear and confusion, or *social*, like Fatima's inability to easily communicate with the people in her neighbourhood. Lower levels of societal participation are correlated with reduced education levels, higher unemployment, lower socio-economic status, limited health literacy, increased social isolation, and lower overall quality of life [4, 9–11].

In this thesis, we have investigated how to support people of low literacy who want to improve their societal participation. This research is a part of COMMIT, a Dutch national public-private ICT research program addressing a range of societal issues such as digital security, healthcare, public safety, and social engagement [12]. Specifically, this work has contributed to the project 'P02: Interaction for Universal Access', that studied the use of ICT and innovative technology to increase social inclusion, self-efficacy, and security awareness for various demographics, including children [13], the elderly [14, 15], and people of low literacy (this thesis). Our goal was to design, develop, and evaluate a software-based learning support system, and to involve low-literate users throughout all steps of the design process. We envisioned a digital learning space in which low-literate learners can individually practice the crucial practical situations in the proper context of use. We believe that practicing the crucial practical situations in a simulated environment will result in highly effective learning that lets low-literate learners gain information and communication skills, improve their self-efficacy about societal participation, and build the motivation and confidence needed to participate in real life. We called the resulting learning support system VESSEL: *a Virtual Environment to Support the Societal participation Education of Low-literates*. The Scenarios section below illustrates how Karel and Fatima could be supported by VESSEL, and what kind of requirements this imposes on the system's design. The rest of this chapter provides background on low literacy in the Netherlands and on computer-based learning support, presents the thesis' research questions, hypotheses, and challenges, and introduces our methodological approach.

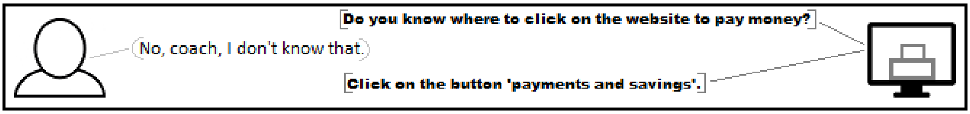
1.2. Scenarios

Karel wants to practice online banking. VESSEL has several scenarios related to online banking. Karel selects the scenario about paying bills with online banking. In this interactive exercise, Karel has to pay for items he bought at a web shop. The screen shows a replica of the online banking page of Karel's bank, which he can interact with using mouse and keyboard. Also visible in the corner of the screen is VESSEL's 'digital coach', a virtual character that will help Karel as needed. The coach introduces itself to Karel, and briefly chats with him about online banking.

Karel initially struggles with the exercise: The online banking website is information-rich and confusing to navigate, and Karel has limited ICT experience. He clicks on various areas of the website, but does not understand what he should do. When Karel does nothing for 20 consecutive seconds, the VESSEL digital coach starts talking to him. First, it asks Karel if he knows what the last button he clicked on means. Karel says no, and the coach explains what it means, in simple Dutch. Karel still does not know what to do, so 20 seconds later, the coach hints that Karel should look for a button about payment. Another 20 seconds later, the coach directly tells Karel to click on the 'Payment and Savings' button. Now Karel understands, and he clicks the button. The coach tells him that this is correct.

The rest of the exercise goes the same way. Sometimes, Karel needs no help to find the next button. Sometimes, the hint that the coach gives is enough. And sometimes, even the direct instruction does not work, as Karel cannot find the right button. In these cases, the coach eventually demonstrates the correct step on-screen. With this help, Karel manages to complete the whole exercise. He is now a little more confident about his ability to pay bills using real online banking.

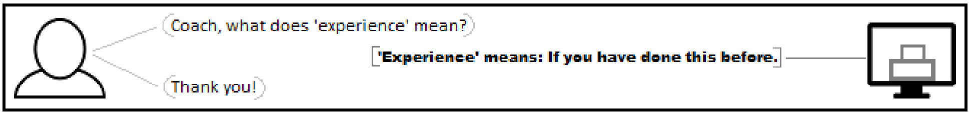
However, he needed a lot of help from the coach. Karel starts the exercise again: If he keeps practicing, eventually, he will be able to do this without any help.



Fatima wants to practice going to the neighbourhood center and talking about volunteer work. There is an exercise about volunteer work in VESSEL: Fatima must talk to a digital character, in Dutch, to do a volunteer work intake meeting. The computer has a microphone she must speak into. The digital coach is present in this exercise as well.

Fatima is a confident learner and dives straight into the exercise. Since she has prior experience with volunteer work, she knows the sort of questions she can expect, and she needs no help with this. Whenever the digital character uses a Dutch word she does not understand, she asks the coach. The coach then explains the difficult word in simple Dutch. This helps Fatima understand the kinds of volunteer work that are available, which sometimes have difficult names. Fatima occasionally gets upset or embarrassed at not knowing something. When this happens, the coach starts talking immediately, telling her that there is nothing to be upset about. Fatima likes that the coach addresses how she is feeling, and this helps her to stay calm.

Fatima manages to complete the exercise quickly. The experience fills her with determination, so she decides to take her practice offline: She heads to the neighbourhood center, ready to try out the same scenario in real life.



These two scenarios show that the virtual coach of VESSEL should be able to provide 'just-in-time' learning support that is adapted to the momentary state and context of the learner. Three kinds of support are needed: Cognitive, affective, and social support. Cognitive support (based on scaffolding) helps the learner understand challenging terms, concepts, and actions, and ensures they can fully finish any exercise; in the scenarios shown earlier, this is seen when the coach asks Karel if he knows what certain buttons mean. Affective support (based on motivational interviewing) helps the learner feel good about their abilities and their progress; in the scenarios, this is seen when the coach tells Fatima she should not be upset when making mistakes. Social support (based on small talk) helps the learner form a bond of trust with the coach; in the scenarios, this is seen when the coach talks to Karel about online banking.

1.3. Background

1.3.1. People of Low Literacy

People of low literacy are defined as adults whose mastery of reading, writing, speaking, and understanding is limited in such a way that they cannot complete the crucial practical situations on their own, meaning they cannot independently participate in society [16]. A sizeable percentage of the population of the Netherlands is low-literate. In 2012, the international *Programme for the International Assessment of Adult Competencies* (PIAAC) survey (which is carried out once every 10 years) reported 1.3 million low-literate people between the ages of 16 and 65 in the Netherlands, around 10% of the Dutch labor force [9, 17]. More recently, the Dutch Court of Audit (*Algemene Rekenkamer*) has expanded the definition to include numeracy issues and adults aged 65+, increasing the total number of low-literate people to 2.5 million [18]. De Greef, Segers and Nijhuis [19] expand on this: Of the 2.5 million Dutch low-literates, 1.8 million are aged between 16 and 65 (a 12% growth from the 2012 PIAAC numbers; a similar increase is reported by [20]). 1.3 million low-literates struggle with either language or both language and maths; 1.5 million struggle with either maths or both. Gubbels et al. [21] report on the result of the 2018 Programme for International Student Assessment (PISA) study, which looked at 15-year-old students, and conclude that 24% of Dutch 15-year-olds possess low language mastery, putting them at risk of low literacy later in life.

Figures 1.1 through 1.4, adapted from de Greef, Segers, & Nijhuis [9] and using the 2012 PIAAC numbers, show how this group is heterogeneous in terms of age, sex, educational history, and cultural background. Fig. 1.1 shows that low literacy increases with age: 5% of people in the age category 16-24 are low-literate, versus 8.1% in the category 25-34, 8.9% in the category 35-44, 14.1% in the category 45-54, and 21.5% in the category 55-65. Fig. 1.2 shows the influence of educational history on low literacy. Of Dutch low-literates, 42.3% have only completed primary school, 38.4% have completed lower vocational high school education, 14.1% have completed higher vocational or scientific high school education, and 2.8% have completed post-high school education. Fig. 1.3 shows that sex is not a determinant for low literacy: The division of men and women across people of low literacy is 47% to 53%. Fig. 1.4 shows the influence of cultural background. This distinguishes between native Dutch people of low literacy, referred to as *NT1* (meaning 'Dutch as a first language', or L1 learners), and non-native Dutch people of low literacy, referred to as *NT2* ('Dutch as a second language', or L2). NT1 citizens make up roughly 65% of the low-literate demographic; their main problems involve reading and writing Dutch, and using ICT correctly. The remaining 35% is NT2 citizens, who primarily struggle with vocabulary and understanding spoken Dutch [9].

1.3.2. Current Learning Approach

In the Netherlands, adult people of low literacy can choose to attend adult language learning classes at Dutch 'regional education center' (*Regionaal Opleidingencentrum, ROC*) schools. Classes focus on language acquisition, societal knowledge,

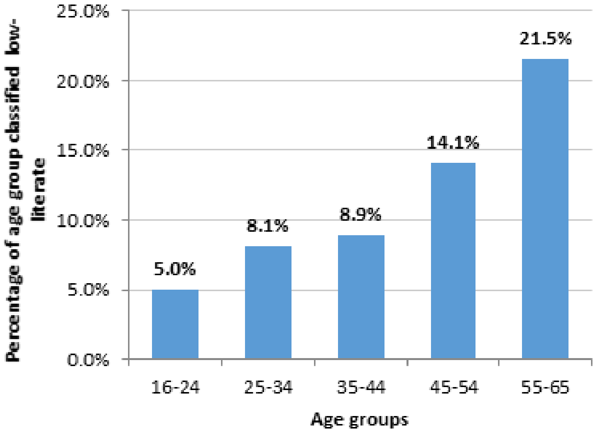


Figure 1.1: Percentage of low-literate people in Dutch labor force, across age

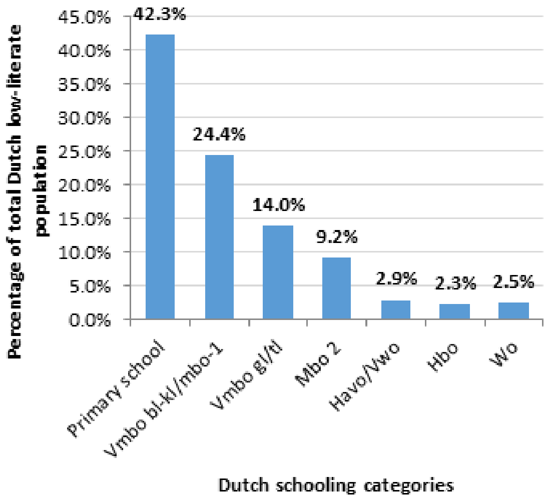


Figure 1.2: Percentage of low-literate people in Dutch labor force, across school history. Categories 'Vmbo bl-kl/mbo-1' and 'Vmbo gl/tl' represent lower vocational high school education. Categories 'Mbo 2' and 'Havo/Vwo' represent higher vocational/scientific high school education. Categories 'Hbo' and 'Wo' represent post-high school education

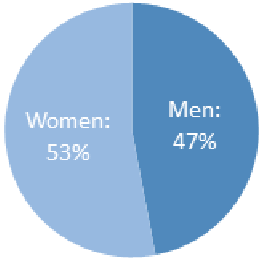


Figure 1.3: Percentage of men and women across low-literate people in Dutch labor force

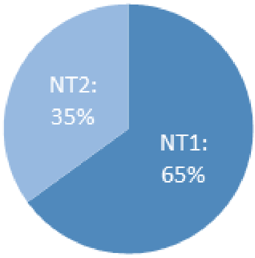


Figure 1.4: Percentage of NT1 and NT2 across low-literate people in Dutch labor force

and training practical skills, all in the context of the crucial practical situations. This is an example of *scenario-based learning* [22], an approach to learning that has roots in constructivism (which treats the learning process as the active construction of knowledge and meaning, cf. [23, 24] and experiential learning (which holds that active learning participation and experience are key to effective learning, cf. [25, 26]). Students discuss crucial practical situation topics in teacher-supervised groups and practice in several ways: They do written exercises from books and worksheets, use learning support software on classroom computers, and roleplay narrative scenarios, either in the classroom or in the appropriate real-life environment [2, 27]. These classes, which until 2015 were a government-mandated part of the education packages offered by the ROCs, are the main source of literacy education for adults; following a policy change in 2015 [28], schools can now freely choose to offer these classes or not, and as a result class provision and attendance have declined [29]. A significant part of the organization of and responsibility for the language learning classes has instead been transferred to a network of external institutions, public libraries, and volunteer work [30, 31].

Classroom learning has attributes that work well with low-literate learners. The presence of the teacher is important: A teacher can provide clear, immediate feedback, and accurately assess the learner's progress and learning needs [32]. Other learners in the class will be low-literate peers, who can help each other with learning problems and provide a safe and welcoming environment. And the classroom as a physical space dedicated to learning is free of distractions and engenders focus. Consequently, de Greef [32] shows that these classes are effective at raising the degree of self-reported educational efficacy (the ability to take part in education), professional efficacy (the ability to take part in the labor market), and social efficacy (the ability to take part in society), which are all aspects of societal participation.

Three areas for improvement can nevertheless be identified. *First*, classroom learning is not accessible to all low-literate people. Learning accessibility can depend on external factors like classroom hours, costs (tuition or travel expenses), and location, but also on the learner's motivation, learning intent, and prior education experience [33–36]. Many low-literates have poor experiences with schooling, which stops them from wanting to go to class [9]. For NT2 learners, the language barrier in the classroom is another impediment: Dutch is the main classroom language, and all books and learning materials are in Dutch. *Second*, classroom learning is difficult to individualize. Large class sizes impede room for individualization as teachers' time and attention are taxed, favoring the use of generalized group lessons over individualized approaches. Individual adaptation would allow learning to better address the varying needs and limitations of the heterogeneous low-literate demographic. Shute & Zapata-Rivera [37] describe three beneficial approaches to learning adaptation: Adaptation to different learner knowledge, skills, and abilities (such as different reading, writing, speaking, and understanding skill levels), adaptation to learner demographic and socio-cultural values (such as the differences between NT1 and NT2 learners), and adaptation to learner emotional states (such as learners that are afraid of going to school). *Third*, not all crucial practical situations that are important to low-literate learners can be practiced in a classroom setting: Certain scenarios may require specialized knowledge or tools, which not every classroom has access to, while others might be socially inappropriate or embarrassing to handle in a group. NT2 learners with different cultural backgrounds may also struggle with established Dutch norms.

1.3.3. Envisioned Computer-Based Learning

We believe that learning can be made more effective for people of low literacy by complementing current classroom learning options with carefully designed learning support software. This is the thought behind VESSEL, the system described in the Scenarios section above. We envision VESSEL as a Virtual Learning Environment (VLE) wherein learners practice with interactive exercises, based on crucial practical situation scenarios and situated in a realistic and contextually correct environment, while the system provides learning support as needed. Learners could practice online banking on a replica online banking website, or practice intake meetings about volunteer work while talking to a digital character in a simulated meeting room. We envision VESSEL as a VLE because VLEs have specific attributes that are useful with respect to supporting low-literate learners. VLEs are a class of software that can contain virtual environments, actors, and objects [38, 39]. *Environments* are (physical or conceptual) spaces for learning to take place in, *actors* are digital characters involved in the learning process, and *objects* are relevant tools or artifacts. Virtual environments can be used to situate scenario-based learning. An example is the aforementioned digital meeting room. We can also imagine a 'digital classroom' that serves as a hub or natural congregation point for learning similar to a normal classroom, while keeping the increased accessibility and adaptability offered by software. Virtual objects can similarly enable or support learning scenarios that are hard to do in real life, while retaining a sense of grounding and physicality. For example, exercises about driving cars, handling dangerous materials, or using expensive work equipment can be simulated in the VLE, combining the practical experience of working with the virtual object with the accessibility and the freedom of consequences of simulation. Finally, virtual actors can fill two roles. On the one hand, users of the VLE can be represented by digital avatars. Here, the social presence and immediate availability of peers and teachers is combined with the accessibility and low social pressure of virtual spaces. On the other hand, a VLE could also incorporate autonomous digital characters, so-called Embodied Conversational Agents (ECAs), into the learning. ECAs are a type of Intelligent Virtual Agents [40], software programs that can interact with other agent programs and with human users, that are 'embodied' as animated characters inside the virtual environment. This affords the ECA new ways of interacting with human users, including body language, gesturing, and non-verbal behaviour [41]. The fact that ECAs have a visual appearance means human users judge them on human-like qualities like similarity and attractiveness [42, 43], and react to ECA behaviours and social cues as if they were human [44–47]. ECAs could help low-literate learners by filling character roles in scenarios, or by taking on the role of a 'digital coach' that provides individualized support to students whenever needed. The potential effectiveness of ECA coaching has been demonstrated in a range of fields and demographics, including computer science education for children [48], health literacy for hospital patients [49, 50], and language learning for second-language learners [51–53].

We think that VESSEL can address the aforementioned shortcomings of classroom-based learning. First, learning software is more accessible, as software use is not necessarily bound to time or location constraints [54–56]. This allows learners to practice at times and places of their choosing, as well as giving them control over the social component of learning. Second, learning software is highly adaptable, making it easier to individualize [22, 57]. Well-designed learning software could incorporate adaptability to user traits, actions, or experiences from the ground up. Finally, learning software affords the practice of scenarios that are inappropriate or impossible to carry out in real life, by simulating them in virtual environments. VLE learning is used for this reason in several domains: For instance, Brinkman, van der Mast, Sandino, Gunawan, & Emmelkamp [58] use virtual scenarios to help people suffering from fear of flying or fear of insects, providing convincing exposure therapy that is cheaper and more fully under the therapist's control than a similar analog scenario would be (cf. [22]). Other examples of virtual environments in learning include training for astronauts [59], city planners [60], emergency first-responders [22], and law enforcement officers [61–63].

Two things are important to note here. One, learning support software used in isolation loses the advantages of classroom learning: There is no teacher to ask for help and no peers to learn with, and learning at home may not have the same sense of focus and urgency that learning in a classroom has. Consequently, we see VESSEL not primarily as a stand-alone solution, but as a complement to existing classroom learning, that learners employ as needed. VESSEL is envisioned as software that learners can use individually, to practice exercises wherever and whenever they want: While this could be at home, it could also be used in a library, or during self-study hours in the classroom. This positions it as a sort of blended learning, a learning approach that combines the advantages of traditional learning and e-learning [64, 65]. Blended learning has been shown to work well with low-literate learners [66, 67].

And two, the envisioned effectiveness of our proposed VESSEL is currently purely hypothetical. To the best of our knowledge, no prior studies have created practical examples or demonstrated the effectiveness of cognitive, affective, and social computer-based learning support for low-literate learners. Consequently, several kinds of important information are still lacking. We do not currently know the exact problems that low-literate people encounter in societal participation or in learning. We do not know what needs and wishes they have with regard to how this learning could be improved. We do not know how a VLE could be designed to meet these needs and wishes, while at the same time remaining accessible, usable, and effective. And we do not know what sort of adaptive support behaviour our digital coach should exhibit to provide effective cognitive, affective, and social support to address these problems and ensure effective learning. These questions cannot be answered *a priori*, but will need to be answered over the course of our work, in collaboration with low-literate learners and subject matter experts.

1.4. Thesis Goals

1.4.1. Challenges

Three main challenges are identified in this thesis. The *first* challenge consists of determining the actual current societal participation problems that low-literate people face, as well as identifying potential solutions. Addressing this challenge will require comprehensive insight into the actor demographic of low-literate people in the Netherlands, the societal participation domain, learning processes, established learning theory, and current technology related to learning support. A mixed-method approach is necessary to address both theoretical and empirical perspectives. This approach will need to be designed carefully: Many user-centered research and design methods and data acquisition techniques assume literacy on the part of the participant, e.g. questionnaires assume participants can read, and interviews assume that participants and interviewers share a spoken language. But these assumptions do not always hold for our low-literate demographic. We require methods that can *a priori* be used by low-literate participants, or that can be adapted to no longer assume or rely on literacy, to ensure that the data we receive will be meaningful.

The *second* challenge concerns designing VESSEL: How do we design effective learning support for low-literate learners, and what attributes and user interaction options must VESSEL have to enable this? We address this challenge by creating a

VESSEL design specification that describes the envisioned system objectives, functional requirements, claims, and use cases. Two questions in particular must be answered. First, VESSEL must be usable by and accessible for low-literate learners. Low-literates in the Netherlands have reduced experience and ability to work with ICT and computers, but are generally not incapable of doing so (cf. [9, 68]). To ensure VESSEL's usability, we must find out what factors govern computer use by low-literates and incorporate these into the design. It also means that learning content (scenarios and exercises) must match low-literate needs, wishes, and learning goals. Currently-available literature provides a starting-off point, but qualitative data such as interviews and workshops will be needed to gain empirical insight. And second, VESSEL should provide effective learning support to low-literate learners, meaning this support should result in high learning effectiveness. To make this possible, we must define what 'learning effectiveness' means in this context and for this demographic: By what metrics do we define whether or not societal participation learning is 'effective' for low-literate learners, and what does this mean for the design of VESSEL?

Finally, the *third* challenge consists of empirically determining the effects of using VESSEL, which we address by evaluating three VESSEL prototypes based on the specification in controlled experiments with low-literate learners. To do this, we must create prototypes that correctly express the specification, and set up effective experiments with low-literate participants, which requires careful selection (and possibly adjustment) of applicable data acquisition methods. The qualitative and quantitative data that results from this must then be incorporated into the specification, iteratively improving the design to better address the aforementioned goals and challenges.

1.4.2. General Research Question

The focus of this thesis is to design VESSEL such that it supports low-literate learners in training practical skills and raising self-efficacy. The general research question that drives this work is:

How can VESSEL support low-literate people in achieving practical exercise success, self-efficacy, and motivation to participate?

1.4.3. Research Questions & Hypotheses

The general research question is decomposed into five design research questions and three hypotheses about learning effectiveness, which we study in five steps. First, we incorporate existing knowledge about the actor demographic of low-literate learners, the domain of societal participation, adult learning, computer-supported learning, and technology into a foundation for the design and evaluation of VESSEL. This is represented in the first research question:

1

Q1. Which operational demands, human factors knowledge, and technologies are important to the design of VESSEL, and which objectives, requirements, and claims can be derived from these?

Second, we acquire empirical insight into the subjective societal participation experiences of people of low literacy: Which cognitive, affective, and social issues low-literate citizens really encounter, how they experience these issues, and what solutions they currently use. The second research question is:

Q2. How can we incorporate the subjective societal participation experiences of low-literates into the design of VESSEL, and which refinements to the VESSEL specification can be derived from this?

Third, we develop a proof-of-concept VESSEL prototype, consisting of information and communication skill exercises and an ECA coach. We explain why we design VESSEL as an ECA coach-supported VLE, create exercises based on the crucial practical situations, and describe how the digital coach can provide cognitive, affective, and social learning support. The corresponding third research question is:

Q3. How can a VESSEL prototype with an ECA coach provide cognitive, affective, and social learning support that meets the operational demands and human factors knowledge?

The effectiveness of the prototype is then tested in a controlled experiment. We predict that a VESSEL prototype using an ECA coach to provide learning support will result in higher learning effectiveness than to a VESSEL prototype without an ECA coach. This forms the first experimental hypothesis:

H1. The VESSEL prototype with ECA coach that provides learning support results in a better learning experience and better learning outcomes than the VESSEL prototype without ECA coach.

Fourth, we formalize VESSEL's cognitive learning support (scaffolding) by creating formal dialogue rules, and by outlining how VESSEL's approach to speech recognition relies on detection of predefined keywords. We describe how the coach tracks the learner's progress in the exercise, how it chooses when to give support and what level of support to use, and how it attunes its support delay to learners' performance in previous exercises, and use this to develop a second prototype. Our fourth research question is:

Q4. How can a VESSEL prototype with an ECA coach provide formal rule-based cognitive learning support that meets the operational demands and human factors knowledge?

We experimentally evaluate the second prototype, predicting that a VESSEL prototype that attunes its support delay to learner performance will result in higher learning effectiveness than a VESSEL prototype that does not. This forms the second experimental hypothesis:

H2. The VESSEL prototype that attunes the support delay to learner performance between exercises results in a better learning experience and better learning outcomes than the VESSEL prototype that does not.

Fifth, we formalize VESSEL's affective learning support (motivational interviewing) by modeling different levels of affective support, defining which emotional states the coach can recognize, and incorporating a heart rate sensor and facial expression detection software to determine learner emotional states. And we formalize VESSEL's small talk-based social support by writing a small talk script. This is used to develop a third prototype, combining the formalized cognitive, affective, and social rules to provide 'full' learning support. Our fifth research question is:

Q5. How can a VESSEL prototype with an ECA coach provide formal rule-based cognitive, affective, and social learning support that meets the operational demands and human factors knowledge?

We predict that a VESSEL prototype that offers full learning support will result in higher learning effectiveness than a VESSEL prototype that only offers cognitive learning support. This forms the third experimental hypothesis:

H3. The VESSEL prototype that provides cognitive, affective, and social learning support results in a better learning experience and better learning outcomes than the VESSEL prototype that provides only cognitive learning support.

1.5. Research Approach

1.5.1. Socio-Cognitive Engineering

The research problem studied in this thesis falls in the category of *design problems* or *design research*, an approach to research that focuses on the adaptation and evaluation of theory in practical applications and contexts [22, 69, 70]. Peeters [22] explains that in design research, the research questions and goals are typically not fully specified from the start, but rather emerge over time or become more accurate as the design process iterates. Consequently, design research requires research methodology that supports rapid iterative work and changing or emerging research questions. We use the Socio-Cognitive Engineering method (SCE), a software design and development method that iterates through three phases: a *foundation* phase wherein relevant operational demands (*contextual information about actors demographics, functional domain, and tasks*), human factors knowledge (*theory relevant to user-system interactions*), and technology drivers and constraints (*both current technology, and envisioned technology to be designed later*) are collected, a *specification* phase wherein data are used to create a design specification consisting of system objectives (*the software's operational or domain goals*), functional requirements (*the software's intended functionality*), claims (*explicit expected outcomes of using the software*), and use cases (*illustrations of how the software is envisioned to be used*), and an *evaluation* phase wherein the design specification is tested with end users, often using prototypes in controlled experiments (Fig. 1.5). The SCE method allows us to combine theoretical and empirical insight in a solid foundation of domain knowledge, and to rapidly and iteratively create and evaluate VESSEL. SCE provides a re-usable design rationale: The focus is not on optimizing specific software, but on developing theory and generalizable knowledge (through software design and evaluation) that can then be applied to other studies or other design projects.

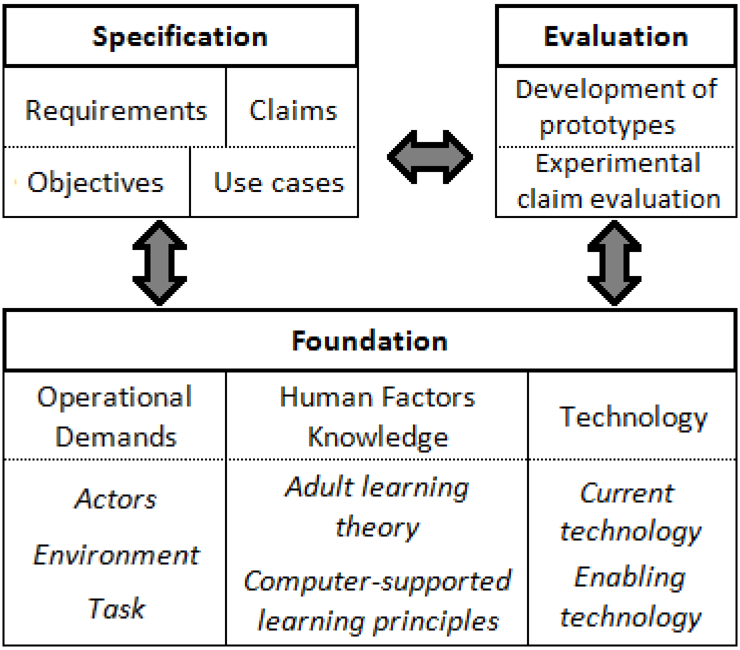


Figure 1.5: Socio-Cognitive Engineering method as used in this thesis

Using SCE as a framework, we identify our needed methods. To build and refine the foundation, we use literature study to gain theoretical insight into relevant operational demands, human factors knowledge, and technology. We also use qualitative data acquisition and analysis methods to gain empirical insight into the needs and wishes of low-literate learners. We combine participant workshops (a hybrid of focus groups and structured data collection exercises, cf. [71] and Cultural Probes (a 'playful' open-ended method that results in rich individual experiences, cf. [72]) to obtain a rich range of data. To transform this foundation into a theoretically and empirically grounded design specification, we combine requirements engineering methods suitable for dealing with theoretical data (cf. [73]) with the Grounded Theory method, which is suitable for dealing with empirical data [74]. To perform evaluations on this specification, we create three VESSEL prototypes, that are used in mixed-method, repeated-measures, within-subjects experiments. When doing requirements engineering/evaluation with a large design specification, it is not always practical to try and evaluate the entire requirements baseline at once [73]. Instead, we create and evaluate multiple prototypes in an iterative process, with each prototype covering a particular subset of VESSEL's full functionality. Stappers et al. [69] describe that the choice of which elements of the larger design to incorporate into prototypes essentially serves as a layer of framing for the study: The elements that you include determine which data you can collect and which questions you can ask and answer, which in turn influences the direction the design research process takes. An example can be seen in this thesis: We choose early on to focus on providing support, while choosing to ignore the element of collaboration. Consequently, support-related findings are incorporated into the foundation of data and used to guide the study, while collaboration plays no meaningful role.

1.5.2. User-Centered Design

We employ a user-centered design approach in this work, by explicitly involving low-literate end users in all steps of the design, development, and evaluation process. We do this to ensure that design choices made throughout the process are thoroughly embedded in the needs, wishes, and values of the target group [75]. This is particularly valuable when the designers and the envisioned end users come from different backgrounds (e.g. socio-economic or cultural differences), as a participatory design approach makes these differences explicit from an early stage [76]. User-centered design can uncover data that would not be accessible through traditional means, and give members of the end-user demographic the power to steer the design/research process in novel directions. As none of the researchers involved in this work have a low-literate background, this approach is important.

We use user-centered design to address three practical challenges related to doing research with people of low literacy. First, we must overcome the problem that the low-literate user demographic is challenging to reach and involve in research, for both practical reasons (e.g. many research methods assume participant literacy) and emotional reasons (e.g. participating in research is seen as frightening or difficult). We address this challenge by involving so-called 'literacy ambassadors': Highly-skilled low-literate people who have been trained to reach out to and encourage other low-literate people [77]. By involving these ambassadors early on, we can ensure that our initial approach is accessible and welcoming, and hopefully make connections with communities of low-literate learners to involve in later work. Second, we must ensure that our research and data collection methods are practically accessible to low-literate people. Cremers et al. [71] show that qualitative data acquisition and analysis methods (e.g. workshops or focus groups) can be very valuable for user-centered design work, as long as the methods are properly calibrated to the target group. We believe the same can be true for quantitative data acquisition, given a careful choice of measures; in either case, we involve low-literate people in the selection and (where necessary) adaption of research methods. While this adaptation is necessary to make the methods practically usable by low-literate participants, we cannot currently gauge the reliability and validity of the adapted methods. Determining these lies outside the scope of this work. Finally, we intend to evaluate VESSEL in controlled end user experiments, where we combine quantitative measures (adapted to the target group where necessary) with qualitative observations to gain comprehensive insight into the effectiveness of our prototypes and design specifications. We involve low-literate people here not only as participants, but also to ensure that our experiments (e.g. our prototypes, our procedures, and our measures) remain practically accessible for low-literates, making changes where needed.

1.6. Thesis Outline

The rest of the thesis proceeds as outlined in Fig. 1.6. In Chapter 2, we address **Q1** by establishing theory about the operational demands of actor demographic, environment, and task, the human factors knowledge about adult learning theory and computer-supported learning principles, and the relevant technologies of existing learning support software and VLEs. We use this to create an initial high-level VESSEL design specification. In Chapter 3, we address **Q2** by incorporating empirical data about the subjective societal participation experiences of low-literates into the design. We use the qualitative research methods 'participant workshop' and 'cultural probe' to gain impressions of the participation experiences and ICT use of low-literate citizens of the Netherlands, and we use the 'Grounded Theory' method to translate these data into the *Societal Participation Experience of Low-Literates* (SPELL) model, which is then used to refine the existing VESSEL specification. In Chapter 4, we address **Q3** by translating our high-level design specification to low-level use cases and a functional proof-of-concept prototype, consisting of four interactive scenario-based exercises and an ECA 'coach' that uses recorded speech utterances to support low-literate learners. We use this prototype to evaluate **H1** in a mixed-method within- and between-subjects experiment. **Q4** is addressed in Chapter 5, where we describe VESSEL's cognitive support functionality in more detail by incorporating scaffolding theory into the foundation. We use a second VESSEL prototype focused on cognitive learning support to evaluate **H2**. We address **Q5** in Chapter 6, where we formalize VESSEL's affective and social support using motivational interviewing and small talk theory and incorporate the 'Shimmer' heartrate sensor and the 'FaceReader' facial expression recognition software into the design to enable autonomous emotion detection. A third prototype is created, which we use to evaluate **H3**. Finally, Chapter 7 presents conclusions for the thesis, discussion, contributions to the scientific and application domains, pitfalls and areas of improvement, and directions for future work.

Chapter 1: Introduction

Chapter 2: Theoretical Foundation

Q1: Which operational demands, human factors knowledge, and technologies are important to the design of VESSEL, and which objectives, requirements, and claims can be derived from this?

Chapter 3: Empirical Refinement

Q2. How can we incorporate the subjective societal participation experiences of low-literates into the design of VESSEL, and which refinements to the VESSEL specification can be derived from this?

Chapter 4: VESSEL Proof-Of-Concept

Q3. How can a VESSEL prototype with an ECA coach provide cognitive, affective, and social learning support as supported by the foundation?

H1. The VESSEL prototype with ECA coach that provides learning support results in a better learning experience and better learning outcomes than the VESSEL prototype without ECA coach.

Chapter 5: Cognitive Learning Support

Q4. How can a VESSEL prototype with an ECA coach provide formal rule-based cognitive learning support as supported by the foundation?

H2. The VESSEL prototype that attunes the support delay to learner performance between exercises results in a better learning experience and better learning outcomes than the VESSEL prototype that does not.

Chapter 6: Affective & Social Learning Support

Q5. How can a VESSEL prototype with an ECA coach provide formal rule-based cognitive, affective, and social learning support as supported by the foundation?

H3. The VESSEL prototype that provides cognitive, affective, and social learning support results in a better learning experience and better learning outcomes than the VESSEL prototype that provides only cognitive learning support.

Chapter 7: Conclusions & Discussion

Figure 1.6: Thesis layout

2

Theoretical Foundation

People of low literacy experience difficulties while participating in society. Learning support software could help alleviate these difficulties. However, there is currently no overview of theoretically and empirically sound requirements for this kind of support. This chapter uses the Socio-Cognitive Engineering method (SCE) to create a requirements baseline for a virtual environment to support the societal participation education of low-literates (VESSEL), based on an analysis of the domain, human factors, and current applications. Four major outcomes are presented. First, a comprehensive overview is collected of the operational demands and human factors knowledge relevant to societal participation learning for low-literate citizens. Second, this overview is translated into a list of eight functional requirements: focused on low-literate learners, set in the context of societal participation, and supported by claims of cognitive, affective, and social benefits to learning. Third, a sample of Dutch societal participation learning support programs are assessed using these requirements, to highlight both current technology best practices and discrepancies between theory and practice. And fourth, Virtual Learning Environment (VLE) technology is suggested as an 'enabling' technology; an overview is shown of how virtual environments, actors, and objects can beneficially enable meeting the requirements baseline. Finally, directions for future study are discussed.

The content of this chapter has been published as: Schouten, D. G. M., Smets, N. J. J. M., Driessen, M., Fuhri, K., Neerincx, M. A., & Cremers, A. H. M. (2016). Requirements for a virtual environment to support the social participation education of low-literates. Universal Access in the Information Society **16**, 3 (2016) [78].

2.1. Introduction

In the Netherlands, societal participation is difficult for low-literate citizens: people dealing with issues stemming from insufficient reading, writing, speaking, and language comprehension skills. De Greef [1] defines societal participation as acting in a society to achieve certain goals. This makes it a socio-behavioral aspect of social inclusion, which is the state of 'being able to take part in society' [79]. Example domains of societal participation include: social interaction with other members of society and formal institutions, societal obligations, self-directed learning and development, and economical and political engagement. Low degrees of participation are associated with unemployment, low socio-economic status, and social isolation [80, 81].

Participating in a modern information society requires that citizens possess the knowledge and the information and communication skills needed to find their own way in society, or know where to go for help [6, 7]. Information and Communication Technology (ICT) skills are increasingly a participation requirement [8]. Because low-literate people have limited language comprehension and communication skills, they are impeded from this. Specific examples of behaviours hindered by low literacy include: Finding work, explaining health concerns to a doctor, socially interacting with peers and neighbours, and using computers and ICT effectively [1, 4, 9, 11].

Three common dimensions of societal participation issues can be seen: cognitive, affective, and social. Cognitive components include the lack of skills, practical knowledge, and experience needed for effective participation. Affective components encompass limited self-efficacy with regard to participation, and feelings of fear, shame, frustration, and stress [11, 82]. And social components consist of relationships with peers, teachers, and other actors. When these relationships are unsupportive, motivation is limited and learning is impeded [1, 10]. Improving the societal participation of low-literate citizens requires learning support that is fine-tuned towards the individual learner's cognitive, affective, and social learning preferences. This can be done, respectively: by connecting learning content to specific problem areas and desired skills, by focusing on the learner's emotional experience, motivation, and self-efficacy, and by forming meaningful connections between learners and the social learning environment.

A range of learning support programs aimed at this area already exists. These programs focus on three topics: 'language learning' trains vocabulary, 'participation skill development' trains the behaviours needed to participate, and 'knowledge of Dutch society' trains Dutch social norms and rules. Training these areas has been shown to significantly improve societal participation behaviour levels [10]. Methods like classroom lessons, roleplaying exercises, book learning, and educational software are used for this training. However, these methods have cognitive, affective, and social drawbacks. Cognitively, lesson plans and materials aimed at larger learner groups are difficult to individualize; this prohibits connecting them to learner skills and interests. Affectively, classroom attendance is difficult for low-literate learners [4]: emotional 'barriers of going to class' can be significant. And socially, mass-produced teaching material can only poorly incorporate the learners' real-life contexts into the learning process. These areas represent room for improvement in the field of societal participation learning support.

Virtual learning environment (VLE) software could provide this contextualized, situated learning support. Virtual environments combine (1) computer-generated spaces and environments, (2) digital actors and characters, and (3) virtual objects and artifacts [38]. Particularly interesting is the fact that VLEs "*...offer the opportunity to simulate a realistic and safe environment for learners to perform specific tasks*" [39, p. 1171]. A realistic VLE designed around societal participation behaviour could help low-literate in several ways. Cognitively, VLEs provide many data visualization options [83], allowing learner skills and limitations to be taken into account more easily. Affectively, the safe and personal nature of VLE learning can reduce the factors of shame and fear of social judgement, eliminating the aforementioned barrier of 'going to class'. And socially, VLEs foster social presence and interaction between students, facilitating group discussion and teamwork and supporting the formation of meaningful social connections. These factors can all engage reticent learners in the learning process.

Supporting the societal participation learning of low-literate citizens through VLE design is the aim of the COMMIT project 'Interaction for Universal Access' [12]. COMMIT is an ICT-focused research programme: in this project, researchers, developers, and consultants investigate how ICT can be used to help low-literate citizens learn about and improve their societal participation. A multidisciplinary approach ensures that all relevant demographical, didactical, and technological angles are taken into account. By acquiring the necessary requirements, and developing the models, methods, and prototypes needed for description and experimental evaluation, the project intends to create a comprehensive specification of the envisioned learning support system '*Virtual Environment to Support the Societal Participation Education of Low-Literates*', or VESSEL.

To achieve this, this study uses the Socio-Cognitive Engineering (SCE) method [84, 85]. This method integrates human factors and technology drivers into an iterative ICT design and development process. The SCE method has three phases. In the *foundation* phase, relevant operational demands, human factors knowledge, and technology are collected into a strong theoretical basis for the design process. In the *specification* phase, this foundation data is used to derive functional requirements. These requirements are contextualized by use cases and scenarios, and justified by verifiable claims. Finally, in the *evaluation* phase, these specification requirements and claims are evaluated. Fig. 2.1 shows a simplified schematic overview of the SCE method, adapted to highlight the focus of this study.

Currently, there is no clear overview of the requirements that VLE software aimed at low-literates should adhere to. This study aims to address this in four ways. First, literature study and domain analysis are used to create a comprehensive overview of the operational demands and human factors knowledge elements of the SCE foundation. Operational demands form a structured overview of the context of use of the envisioned software [86]: the actor demographics of low-literate learners, a description of 'societal participation behaviour', and the cognitive, affective, and social processes important to the task of societal participation learning.

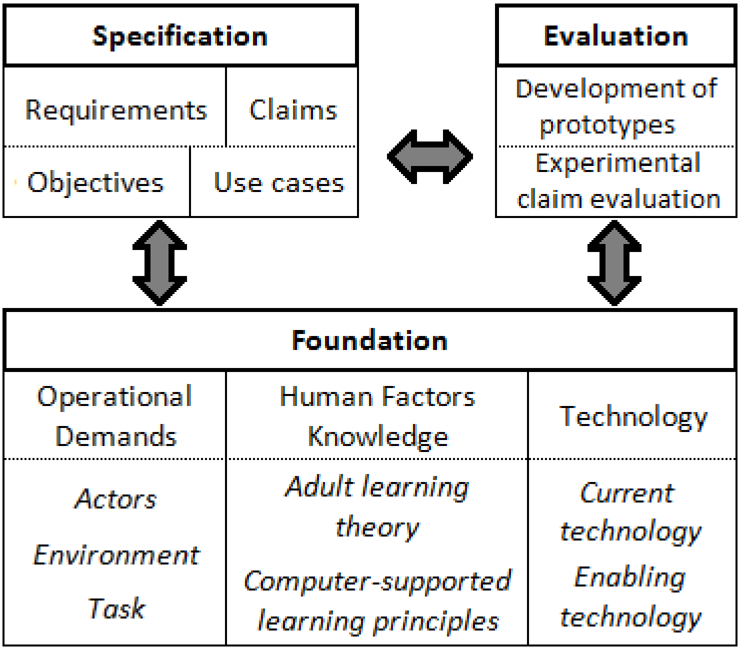


Figure 2.1: Socio-Cognitive Engineering method. Phases are shown: the foundation phase (lower box), the specification phase (top left box), and the evaluation phase (top right box) [84, 85]

And human factors knowledge presents insight into adult learning theory and principles of computer-supported learning. Second, the established foundation is used to derive an initial list of functional requirements for the proposed VESSEL system. Cognitive, affective, and social claims of learning benefit are also derived. Third, a selection of existing learning support programs is assessed using the derived requirements. The goal of this assessment is to explore which requirements are and are not met in daily practice. This highlights discrepancies between theoretical and practical importance, and collects practical examples of requirement implementation. And fourth, VLE technology is proposed as a possible 'enabling' technology for meeting the VESSEL specification. Virtual environments, actors, and objects all have particular characteristics that make it easier to effectively implement most of the requirements. In this way, this study aims to answer the following questions:

1. Which attributes of the actor demographics, societal participation behaviours, learning processes, adult learning theories, and computer-supported learning principles, are relevant for the design of VESSEL?

2. How can these attributes be used to adapt learning support to the skills and characteristics of individual users? Which functional requirements and claims can be derived?

3. Which of these functional requirements are met by current learning support

programs? And which ones are not? And what lessons can be learned from this?

4. How can these requirements and lessons be used in the design of VESSEL? And how can the use of virtual environment technology help with this?

The structure of this paper is as follows. Section 2.2 forms the operational demands part of the SCE foundation, describing the actor demographic of low-literate people, defining the conceptual environment of 'societal participation behaviour', and providing insight into the cognitive, affective, and social processes underlying the core task of learning. Section 2.3 forms the human factors knowledge part, investigating frameworks of adult education and ICT learning principles. Section 2.4 makes use of the current foundation to derive a specification for VESSEL: a list of functional requirements and claims. Section 2.5 forms the technology part, respectively listing examples of current societal participation learning support programs, assessing these programs on the basis of the functional requirements, and expanding on the unique attributes of virtual environment software. Section 2.6 forms the conclusion.

2.2. Operational Demands

Operational demands comprise three main categories: actors, environment, and task (i.e. what user group will the design be aimed at, what environment will the design be used in, and what task is the design intended for). First, the *actor* demographic of 'low-literate Dutch people' is explored, and further defined in terms of learner profiles. Second, the *environment* in which societal participation takes place is described. And third, the *task* of societal participation learning is defined by describing the cognitive, affective, and social processes involved.

2.2.1. Actors: Low-Literate Citizens

This section describes the actor demographic of low-literate Dutch citizens. After defining the concept of 'low literacy' in functional terms, the corresponding demographic information is presented. Building on this, five 'learner profiles' are introduced, and their relevance to software design is explained.

Literacy

Buisman & Houtkoop [4] use results from the 2012 Programme for International Assessment of Adult Competencies to define the *core skills of societal participation*. These core skills include literacy, mathematics ability, and general problem solving. *Literacy* is the ability to comprehend, process, and make use of information. The Organization for Economic Co-operation and Development (OECD) defines literacy as "*the ability to understand and employ printed information in daily activities, at home, at work and in the community – to achieve one's goals, and to develop one's knowledge and potential*" [16, p. X, *sic*]. Literacy is an important determinant for successful societal participation, particularly in modern information societies [5, 8]. Anyone whose mastery of this core skill is too low to allow them to act and live as an independent citizen is considered low-literate.

Demographic Information

Currently, about 1.3 million people between the ages of 16 and 65 living in the Netherlands are low-literate [9]. This works out to around 10% of the labor force. This percentage has remained stable for the past two decades, and is projected to persist until at least 2020 [68]. The collection of people of low literacy living in the Netherlands is not homogeneous. Based on language background, two broad groups have been defined: 'NT1' and 'NT2'. 'NT1' refers to native Dutch people of low literacy, who are said to be learning 'Dutch as a first language'. 'NT2' refers to Dutch citizens with a non-Dutch mother tongue, who are said to be learning 'Dutch as a second language'. Included in this second group are both low-educated first-generation migrant citizens, who are functionally low-literate in both their mother tongue and in Dutch, and second-generation migrants for whom Dutch is not a mother tongue, and who have often seen little writing in their upbringing. Crucially, this group does *not* contain migrant citizens who are functionally literate in their mother tongue but not in Dutch, as their particular issues fall in the field of second language acquisition.

In theory, the two groups are different enough in skills, problems, and context to merit individual study. In practice, demographic studies report large overlaps in terms of literacy-related issues. While significant differences between the experiences of NT1 and NT2 learners exist, the shared issue of low literacy suggests strong, meaningful commonalities in problem areas, solution directions, and support possibilities. As such, the term 'low-literate' is used here to encompass all literacy-impaired citizens of the Netherlands, regardless of background.

De Greef, Segers, & Nijhuis [9] further describe this demographic. Low literacy increases with age: while only 5% of citizens aged 16-24 is low-literate, 21.5% of citizens aged 55-65 is. Education levels among low-literates tend to be low, with as many as 42% not surpassing primary school levels. Low-literacy affects men and women almost equally. Finally, roughly two-thirds of low-literates are native citizens, with the rest split over first- and second-generation migrants. These statistics provide design and study guidelines: low-literate citizens are likely to be older men and women, more often native than non-native, and poorly educated.

Low-literates are less likely to work with computers and technology than people of higher literacy. This suggests a possible 'digital gap': a divide between the high-literate people capable of working with modern computer technology, and the low-literate people incapable of doing so [87, 88]. This would have negative repercussions for using software to provide learning support. However, these fears may be unfounded. Houtkoop et al. [68] report relatively high technology usage statistics among Dutch low-literate citizens. Nine out of ten low-literates have home access to a personal computer, and access to the internet. Three quarters of low-literates have some significant computer experience, and almost half regularly use a computer at work. In the Netherlands, only one in five low-literates is seriously lacking in computer skills [9]. Furthermore, significant numbers of Dutch low-literates exhibit personal interest in computer skills learning, and they expect that computer skills learning will become relevant to their situation in the near future [68]. Correspondingly, while care is still needed, there is no indication that learning support software would be significantly less effective for the majority of low-literate learners than it would be for literate learners.

Learner Profiles

Based on studies of language learning class attendants, Kurvers, Dalderop, & Stockmann [89] have derived five low-literate learner profiles. These profiles divide the low-literate demographic in strata: This is based on language background, educational history, current literacy level (based on the Common European Framework of Reference for Languages, cf. [90]), needed improvements, and potential complicating factors. These profiles (detailed in Table 2.1) are useful for study and software design, for example in classification and prediction. Not all learner profiles are equally interesting in the context of societal participation software design. For both profiles 1 and 5, a software-based solution seems relatively ineffective. Learners in profile 1 are at a high enough level of skill and self-direction that software learning support will not provide much benefit. And the 'difficult' learners in profile 5 are too low-skilled and literacy-deficient for a computer-based solution to be functionally applicable. Profiles 2, 3, and 4 could still benefit from societal participation learning support. Consequently, these three profiles are used throughout the design process.

2.2.2. Environment: Societal Participation Behaviour

As noted by de Greef [1] and Schouten [91], societal participation is expressed through the goal-directed social behaviours of citizens. Societal participation is behaviour. Similarly, 'improving societal participation' can be understood to mean 'learning to better perform goal-directed social behaviours in a societal context'.

The three key elements of societal participation behaviour are *language*, *societal knowledge*, and *participation skills*. The 'language' aspect refers to the ability to effectively communicate in and participate in modern society. According to Breen & Candlin, language learning involves "*...learning how to communicate as a member of a particular socio-cultural group*" [92, p. 90]. Particularly for low-literate second language learners, limited vocabulary is a major participation impediment [89]. The 'societal knowledge' aspect indicates knowledge of how to act in society. Low-literate second language learners often follow cultural norms and assumed rules, which may be spread around by word of mouth and experience, instead of formal written rules. For native low-literate learners, problems occur in situations where the required information and communication skills are too complex. Finally, the 'participation skills' aspect describes the functional skills needed to be successful at participation behaviour.

Schouten [91] has created a model that uses the dimensions of skill and context to map out the societal participation behaviour domain (see Fig. 2.2). In this model, the skill dimension ranks behaviour on varying degrees of *information* skill (the ability to comprehend and process information) and/or *communication* skill (the ability to communicate with others). And the context dimension describes the kind of social setting the societal participation behaviour takes place in: Either more *formal*, characterized by rigid structures and a less personal atmosphere, or more

	Language background	Educational history	Current literacy levels	Complications	Needed improvements
Profile 1: The advanced learner	57% NT1 43% NT2	9 years of formal schooling on average. 70% regular schooling. 30% special schooling.	Reading: A2 to B1 Writing: A2 to B1 Speaking: B1	-	Writing
Profile 2: The average learner	57% NT1 43% NT2	9 years of formal schooling on average. 70% regular schooling. 30% special schooling.	Reading: A2 Writing: A2 Speaking: A2	-	Reading, writing, occasionally speaking
Profile 3: The non-native learner	0% NT1 100% NT2	9 years of formal schooling on average. 70% regular schooling. 30% special schooling.	Reading: A2 to B1 Writing: A2 to B1 Speaking: A1 to B1	Possess limited vocabulary, which limits reading, writing and speaking.	Speaking skill and vocabulary
Profile 4: The low-skill native Dutch learner	100% NT1 0% NT2	Almost entirely special schooling. Indeterminate length.	Reading: A1 to A2 Writing: A1 to A2 Speaking: C1 (native)	Dyslexia is either suspected or confirmed in most of these learners.	Reading, writing, functional literacy
Profile 5: The difficult learner	Unspecified mix of NT1 and NT2	Unspecified. Very little.	Reading: A1 or below Writing: A1 or below Speaking: A1 or below	Little to no skill at learning: Learning disabilities suspected very limited schooling.	Dependent on individual circumstances

Table 2.1: Low-literate learner profiles. Columns, from left to right: Profile description. NT1/NT2 distribution. Average educational. Learner literacy levels (using CEFR levels). Additional common complications. Common needed improvements. Data from [89].

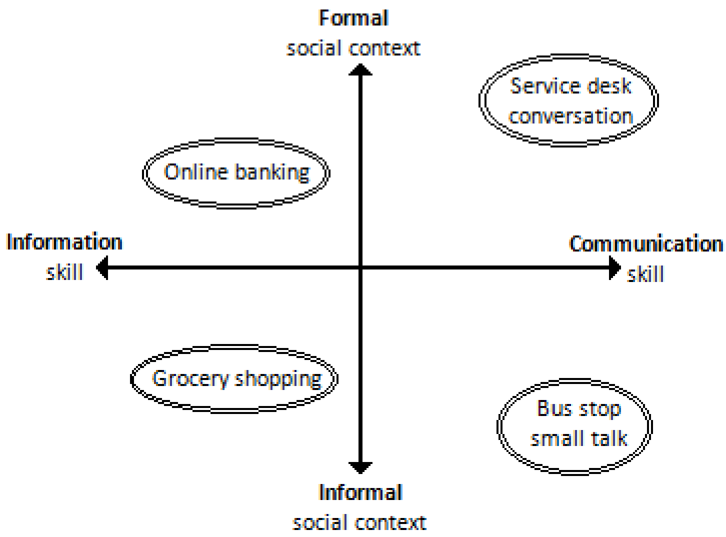


Figure 2.2: Societal participation behaviour model. Based on [91]

informal, characterized by a less imposing and more open-ended nature. Fig. 2.2 also shows examples of societal participation behaviours that low-literate citizens have been found to struggle with.

2.2.3. Task: Societal Participation Learning

This section expands on the task of learning, which makes up the core of 'societal participation learning'. Three categories of processes that influence the effectiveness of the learning process have been defined: cognitive processes, affective or emotional processes, and social processes.

Cognitive Processes

Cognitive processes refer to those processes that influence the rational, cognitive relation between learner and learning. Cognitive processes involve elements like reading and writing skill, memory, domain-specific knowledge, learning aptitude, prior experience, and task difficulty. The influence that cognitive and meta-cognitive strategies like scaffolding (cf. [93]), planning, organizing, and monitoring have on learning effectiveness and academic outcome is well-documented [94–96]. Cognitive processes regulate the balance between learning difficulty and learner skill, forming a strong determinant for learning success [97]. *Transfer of learning*, the degree to which the learned material transfers to the learner's daily life, is a particularly important cognitive measure of learning effectiveness [1, 97], especially in behaviour-oriented domains like societal participation.

2

Affective Processes

Affective processes refer to those processes that influence the affective, emotional relation between learner and learning. Affective processes involve the learner's emotional state, their self-image as a capable learner, and their feelings about the learning process. The important role of affective processes can be understood by looking at *self-efficacy*. First described by Bandura [98, 99], self-efficacy refers to an individual's task-specific judgment of their own capabilities. In a wide range of fields, self-efficacy is a powerful predictor of behavioral intent, motivation, and performance. These fields include academic achievement [16, 98, 100–102], reading and writing [103], computer use and adoption in general [87, 88], and computer use specifically for learning [104, 105]. Low self-efficacy is a significant factor in explaining the low societal participation of low-literate citizens. Van Linden & Cremers [82] and Mertens & van het Zwet [11] demonstrate the harmful nature of low 'societal participation' self-efficacy. Cremers, de Jong, & van Balken [106] show the use of Automated Teller Machines (ATMs) as a concrete example of the practical issues low self-efficacy can cause. Schouten [107] further claims that low self-efficacy with regard to learning and academic achievement inhibits learning behaviour for low-literate learners (cf. [108]), and that low self-efficacy about technology literacy and computer use impedes the effectiveness of learning support software.

Next to self-efficacy, additional affective roles are played by emotions like fear and shame. The deleterious effect of fear on academic engagement and success is well-known [109, 110]. And shame about low-literacy has been shown to inhibit societal participation, learning engagement, and learning effectiveness in low-literate citizens [111, 112].

Social Processes

Social processes refer to those processes that influence the social, environmental relation between learner and learning. Social processes relate to other humans involved in learning, like peers and teachers, as well as to the environments in which learning and practice take place. These processes often reflect on the learner's judgement of the learning process itself: negative attitudes towards learning can often be traced back to poor formative classroom experiences (cf. [108]). In this light, *motivation to act* is an important social process (cf. [113]). The effects of motivation on behaviour choice and persistence [109] are well-known, and motivation has been related to learning behaviour [95, 114], academic success [115], writing [116], and computer technology use [117]. The social aspect of motivation can be seen in the influence of encouragement or discouragement from social peers. Motivation to *get started* in the learning process has particularly strong social components: positive or negative first interaction with new teachers and learning peers can determine later learning intention, either guiding new learners into the process gently or scaring them off altogether.

Motivation to go to class is supplanted by *interest in* and *engagement with* the learning process. Interest and engagement have been linked to academic success. Facer et al. [118] show that engaged, active students are interested in the learning process, while unengaged students are passive and resistant to learning. Garcia

& Pintrich [95] list personal interest and 'a sense of achieving personal goals' as a subset of academic performance predictors. Levy [119] additionally suggests that personal interest in the learning environment works as a predictor for *learner retention*: learners who feel satisfied with their learning environment are less likely to drop out. Both Parker [120] and Schunk [102] suggest that learner locus of control factors into retention as well. Learners who feel like they lack choice in working on a task are less likely to keep going than learners who feel like they are in control.

2.3. Human Factors Knowledge

In this section, human factors knowledge relevant to the design of societal participation learning support software is presented. First, adult learning theories are investigated, in order to present an overview of the current didactical state of affairs. *Adult* learning theory is specifically selected because the low-literate target demographic, outlined in section 2.2.1, covers the 16-65 age range. And second, computer-supported learning principles are investigated, in order to clarify in what ways the use of computer technology can help. This information has been collected by way of literature research, and is presented in the following sections.

2.3.1. Adult Learning Theory

To date, there is no single unifying theory of adult education. A range of theories observes adult education from different perspectives and reaches different conclusions and recommendations (e.g. research on learning styles and preferences (cf. [25, 121]). Three didactical method frameworks seem particularly interesting: Andragogy, transformative learning, and constructivism [6, 54, 96, 122].

Andragogy postulates that adult learners possess several characteristics, developed during adolescence or adulthood, that explain why adults learn in ways significantly different from children [123, 124]. Six learner characteristics have been identified [125]: adults are self-directed learners, driven by real-life problems, internal motivations, and societal roles and demands, who want to know why they should learn anything they are told to learn and possess accumulated life experience to draw on. These characteristics seem to offer clear and simple guidelines for adult education (but see [108] for a rebuttal). Andragogy is interesting for the notion that not all learners have similar learning styles and preferences. Because of significant differences between low-literate individuals and between learner groups, societal participation learning should never be offered in one singular way. Successful societal participation learning support must be **adjustable and adaptable to individual needs and preferences**, in order to take into account different learning styles and preferences, skill levels, and difficulty curves.

Transformative leaning is a type of learning that involves altering frames of reference. Mezirow describes transformative learning as "*...the process by which we transform our taken-for-granted frames of reference*" [126, p. 7]. Illeris [127] contrasts transformative learning with assimilative learning (which keeps frames of reference intact) and accommodative learning (which involves restructuring frames of reference). Rather than dealing with knowledge acquisition, transformative learning is about the learner evaluating and changing their views and assumptions on the world. The transformative learning viewpoint is valuable for highlighting the importance of a careful, sensitive approach. Societal participation learning support needs to recognize the volatility and negative affect associated with societal participation learning, and attempt to **employ sensitivity** to defuse or prevent it. Examples of sensitivity include: using non-confrontational language and learning examples, respecting users' desire for privacy and anonymity, and demonstrating situational and cultural awareness [57].

Constructivism sees 'learning' as the active construction of knowledge and meaning. Learners engage in this construction process through interaction with other learners, and with their own environment, experiences, and ideas [23, 24]. In recent decades, the focus has partially shifted to the collaborative and social dimensions of learning. This particular view on constructivism is called *social constructivism* [128]. One social constructivist notion is that of authentic or situated learning, also known as *situated cognition*. The main concepts of situated cognition are that all learning takes place in a certain context, and that the context in which knowledge is presented is as much a part of the learning as the knowledge itself. This context includes the physical location learning takes place in, the tools used and their method of use, and the social interaction with other people. Hansman claims that "*The core idea in situated cognition is that learning is inherently social in nature.*" [129, p. 45]. And Brown, Collins, & Duguid [130] and Lave & Wenger [131] have emphasized the role of social interaction in learning, and investigated the potential benefits of situated cognition and situated learning. Studies have suggested that affective dimensions and emotions form an important situated element in the context of learning as well [6, 132, 133]. As societal participation behaviour is intrinsically situated, it follows that societal participation learning should also be. Because the goal of societal participation learning is to teach real, applicable knowledge, learning should be situated in an environment as close to the learners' real-life environment as possible. Societal participation learning support must be able to **situate the learning process** in the context of events, locations, behaviours, and actors that learners are likely to encounter in their day-to-day lives.

Another notion of social constructivism is its focus on social, collaborative learning. In the context of low-literate societal participation learning, both teacher-directed learning and peer-interaction learning have their place. Consequently, societal participation learning support should provide various methods of **social interaction and collaboration in learning**, in either fully digital or blended forms, and encourage the application of these possibilities in the learning process. Many existing *e-learning* theories and applications posit this social constructivist view of learning, where personal meaning-making and the social influences of peers and teachers shape the learning process and outcomes [134–136]. Initially referring to computer- and ICT-supported distance learning and computer conferencing, e-learning now encompasses a style of learning that focuses on learner collaboration, technology-supported communication methods, and the formation of digital communities of inquiry [54]. One of e-learning's typical features is the elimination of the classroom as the necessary physical hub for learning. Modern communication technology has all but eliminated distance as a critical factor, allowing learners and teachers in various locations to engage in joint learning processes. The utility and added value of face-to-face contact are still acknowledged, however [54]: so-called *blended* approaches, which mix elements of classroom learning and e-learning [65] are becoming increasingly prevalent. Steehouder & Tijssen [67] and Driessen et al. [66] report on the effectiveness of using blended learning with low-literate learners.

In conclusion, four concepts have been identified as being important for the design of societal participation learning support: adaptability to learning styles and preferences, sensitivity, situatedness, and collaboration.

2.3.2. Computer-Supported Learning Principles

ICT has always held promise in the field of education. However, its effectiveness seems highly contingent on proper introduction and use. According to Cuban [137], forcing ICT measures on unwilling teachers leads them to use the computer like a 'replacement typewriter'. And Sansone et al. [110] report that unsupported at-home students display both lower motivation and poorer results than on-campus students. This section lists a number of ICT-related learning principles, derived and adapted from Richards [56]' topology of meaningful ICT learning activities. These principles are: the provision and dissemination of information, the possibility and facilitation of worldwide communication, the element of interactivity, and gaming principles.

Provision of information refers to the fact that ICT learning offers a wide range of media and information types. Video, audio, and written text can all be offered in conjunction. These possibilities make it easier to adapt elements of a learning application to individual learners' preferences. Low-literate learners, for example, could benefit from an implementation focusing on audio and video, from supporting material offered in multiple languages and at different language levels, and from language-and culture-specific elements such as avatar ethnicity and dress style. Societal participation learning support should adapt to the needs and wishes of the learners as much as possible; modal adaptability to individual user preferences is an important example of this. Used correctly, **multimodality** can remove significant barriers to entry and learning progress.

Worldwide communication is one of the cornerstones of ICT, allowing teachers and students to stay in contact beyond normal classroom hours. This opens up venues for directed, personalized support. This is vital in ICT-learning: Nielson [55] reports high rates of failure in at-home language learning without proper support. Furthermore, it can be argued that the near-total dissemination of ICT use in information societies [138] has turned 'the proper usage of ICT communication tools' into an important societal participation skill in its own right. Given that low-literate citizens often possess reduced ICT skill levels, ICT-based societal participation learning support could serve a dual purpose in acquainting learners with ICT practices and behavioral norms. The findings by Nielson [55] and Driessen et al. [138] show that proper learning support is vital in societal participation learning. The use of ICT in this support seems a natural fit, as instantaneous, ubiquitous, and easily adaptable communication is one of its hallmarks. As Schouten [91] argues that poor learning experiences and dissatisfaction with the learning process form major participation hurdles for low-literate learners, societal participation learning support should make use of this ICT-related **learning support** to improve the learning experience wherever possible.

Interactivity links ICT learning to experiential learning [25, 26]. Barak [57] divides learning into four aspects – contextual, active, social and reflective – and reports that the use of ICT enhances the contextual and active parts of learning. This implies that the interactivity of ICT learning applications is tied to both experiential learning and situated cognition, as adapting to different learning styles and preferences and to different contexts is easier in ICT than in more traditional teaching methods. For optimal learning results, societal participation learning support should actively involve users in the learning process. Good use of **interactivity** could be beneficial for low-literate societal participation learners.

Digital gaming is increasingly seen as a form of experiential learning [139–141]. Doshi [142] claims that using gaming to teach skills allows students to fit otherwise abstract concepts into their daily lives. Ke [143] suggests that educational games offer four potential benefits to learning: Games are a conduit to experiential learning, games create and enhance engagement in students, games promote cooperation, and games could help students in digesting complex subject matter. Evidence supports some of these claims. Studies show that games can induce engagement and immersion [144], and flow and fiero [59]. Both Dickey [145] and Warren et al. [116] report that games can increase intrinsic motivation in players. Rieber [146] claims that students view gaming as 'play', but regular learning as 'work'. Gaming has also been shown to promote cooperation among school children [118]. Finally, Kriz posits that "*...gaming simulation represents a form of cooperative learning through teamwork.*" [147, p. 506]. Gaming principles and gamification can be used to enhance the effectiveness and the affective and social experience of learning. Astell et al. [148] show that video games can be used with older adults with cognitive impairments, both to help them learn and to engage and entertain them. Similar expectations can be held about the use of gaming elements with low-literate learners. Games, and other forms of 'playful' learning, could also go a long way towards improving negative attitudes with regard to societal participation and to learning (cf. [91]).

In conclusion, four concepts have been identified as being important for the design of societal participation learning support: multimodality, learning support, interactivity, and gaming principles.

2.4. Specification: Requirements

In this section, the theory presented by the human factors knowledge and the context of use sketched by the operational demands are brought together to specify the requirements for the proposed VESSEL system. Claims that underlie these requirements are derived as well. These claims make explicit in what ways each requirement is expected to influence the cognitive, affective, and social processes related to learning. The requirements represent *what* learning support software should do, and the claims represent *why* this should be done (thus providing the design rationale). This makes up the specification stage of the SCE method [84, 85] (see Fig. 2.1), and forms the specification for VESSEL. Requirements are summarized in the list below. A schematic overview is also shown in Table 2.3.

R1. Adaptability: A societal participation learning support program should offer and support different learning styles and preferences and different difficulty levels.

- *Cognitive Claim*: Catering to different learning styles and preferences will allow individual low-literate users to select those styles and difficulty levels that work best for them. This will match learning content and process to user experiences and aptitudes, and improve learning effectiveness over a one-approach-fits-all system.

- *Affective Claim*: Allowing individual users to indicate their own learning styles, preferences, and difficulty levels provides the user with a modicum of power and influence over their own learning experience. This will result in a more positive user experience.

R2. Sensitivity: A societal participation learning support program should use non-confrontational language and content, demonstrate cultural awareness, and take existing emotional issues with regard to literacy and societal participation into account.

- *Affective Claim*: Demonstrating awareness of individual problems and difficulties, and showing a willingness to take these elements into account in the learning process, will give users a sense of being respected and of being listened to. This will result in a more positive user experience.

- *Social Claim*: Learners who experience that their individual problems and difficulties are being respected by teachers, peers, and learning materials will be more likely to continue learning. This will improve learner retention.

R3. Situatedness: A societal participation learning support program should use learning materials and contexts that are closely related to the learner's physical environment and real-life experiences.

- *Cognitive Claim*: Situating learning and support in the context of real-life situations will help in transferring useful, applicable skills and experiences. This match between user experiences and presented learning experiences will improve transfer of learning.

- *Affective Claim*: Particularly for those low-literate learners with low learning self-efficacy, placing the learning content in a well-known personal context will reduce barriers of fear and uncertainty.

- *Social Claim*: The use of recognizable environments, actors, and skills makes learning more engaging and immersive for learners. This will improve learner retention.

R4. Collaboration: A societal participation learning support program should have systems in place that enable, support, and foster social interaction and collaboration in learning.

- *Cognitive Claim*: Learning about societal participation in a social and collaborative setting will create scenarios and produce knowledge and experience that are more closely applicable to real-life participation. This will enhance learning applicability and transfer of learning.
- *Social Claim*: The presence of peers and teacher support mixed with the privacy and safety of technology-supported learning will reduce the barriers low-literate learners experience in starting with and persisting in learning endeavors. This will improve motivation to start learning and learner retention.

R5. Multimodality: A societal participation learning support program should employ multimodality, offering content in multiple concurrent ways.

- *Cognitive Claim*: Given that reading is a particular difficulty for low-literates, using presentation modalities other than text is preferred. However, differences between low-literates make any monolithic approach untenable. Engaging multiple modalities ensures that all low-literate learners have some preferred way of accessing learning. This will improve learning effectiveness over a system that does not use multimodality.
- *Affective Claim*: Self-conscious low-literates learners are often worried that reading skill will be required to participate in learning. Knowing that materials are offered in some other format than written text alleviates this fear, removing a significant barrier of stress and anticipation for these learners.

R6. Support: A societal participation learning support program should possess built-in support options.

- *Cognitive Claim*: Good support options will aid users in understanding the learning material better. This will improve learning comprehension and effectiveness.
- *Affective Claim*: Low-literate learners value the idea of being supported. A desire for individualized, personal contact with supporting experts and peers has been reported by multiple studies with various demographics. (cf. [149]) on video support for caregivers) Providing proper support will set users at ease while using the learning software. This will increase their self-efficacy, both with regard to societal participation and with regard to learning itself.
- *Social Claim*: A learner who receives support when they want it or need it will feel supported. This will lead to higher learner retention: learners are more motivated to continue when they know that help is available.

R7. Interactivity: A societal participation learning support program should employ real interactivity in offering content.

- *Cognitive Claim*: Interactivity in learning behaviour is almost ubiquitous at this point in time, and the educational benefits of learning-by-doing and scenario-based learning are well-documented, particularly with regard to learning transfer [97, 143].
- *Social Claim*: Negative attitudes towards the learning process can often be traced back to poor prior classroom experiences. Interactive learning will help alleviate these negative attitudes, by engaging low-literate learners more in the learning process.

R8. Gaming Principles: A societal participation learning support program should use elements and principles of interactive gaming.

- *Affective Claim*: Using gaming principles will mitigate the negative affective view low-literate learners tend to have about learning in general (cf. [91]). This will result in a more positive learning experience.
- *Social Claim*: Gamification is often essentially an attempt to make learning more 'fun'. Doing this will improve engagement with and immersion in learning, and foster motivation to persist.

2.5. Technology

In this section, an overview of technology relevant to the design of societal participation learning support is presented. First, the current state of technology regarding Dutch societal participation learning support programs is investigated. Examples of learning support software are collected and described. Second, these software examples are assessed on the basis of the VESSEL requirements drafted in section 2.4. This assessment intends to highlight which requirements are commonly met in current practice, and which ones are not. And third, the enabling role of VLE technology is described. An overview is presented of how the core VLE attributes 'environments', 'actors', and 'objects' enable and support the effective implementation of many of the VESSEL requirements.

2.5.1. Current Technology

In this section, an overview is shown of currently existing learning support software programs that aim to improve the societal participation of low-literate learners. This overview was created by searching for software that adheres to three characteristics. First, the software must be intended to offer **learning** to students about the topic at hand. Second, the software must be intended for use by **low-literate learners**. And third, the software's subject matter must be **societal participation learning**. This means that it must focus on **language acquisition**, **societal participation skills**, and/or **knowledge of Dutch society**.

Six software packages met these characteristics. These software packages were all integration training courses, aimed primarily at low-literate second language learners. One package also included material for native low-literate language learners. No packages that focused only on native low-literate language learners were found. None of the software packages were stand-alone products. Rather, each software example was part of a larger integration learning program. The software was designed to be used in concert with other materials: books, worksheets, classes, and practical exercises.

The following packages were found: *EHBN*, *EVT.nl*, *Naar Nederland*, *Thuis in Nederlands*, *IJsbreker Plus*, and *NL247*. Summaries of these software packages, including description, publisher, production year, and learning focus, have been included in Appendix A for purposes of study reproducibility.

2.5.2. Assessment Of Current Practice

This section provides an assessment of the described learning software 'in isolation'. It should be noted that the software applications are part of larger, multifaceted educational programs that are not evaluated. The interest here is in the functionality of the software only: the intent is to create an overview of which of the theory-based requirements and claims are met in current software design practice, and which ones are not. The former will provide insight into practical, effective ways of operationalizing these requirements. The latter will highlight interesting areas for future study and development.

Table 2.2 shows that the various requirements are reflected in software design at different levels of frequency. While the demands for sensitivity, interactivity, and multimodality are fulfilled quite often, examples of adaptability, collaboration, gamification, and support are rare. The sensitivity (**R2**) requirement can be found in all evaluated examples. The implementation of sensitivity can be seen in careful word choices, and in appreciation for cultural differences. Even brief study provides many examples to emulate in future design.

As mentioned in section 2.3, interactivity (**R7**) and multimodality (**R5**) are cornerstones of ICT technology and design. The combined cognitive benefits of ease-of-access and improved learning transfer greatly enhance learning effectiveness. And the affective benefits of lowering barriers of stress and anticipation, offered by multimodality, are almost necessary while designing for low-literate learners. Again, all studied software packages make good use of these possibilities: audio and video supplement text, and exercises and lessons are often interactive. Examples can be seen in *IJsbreker Plus*' audio-supported multiple-choice questions, in *ETV.nl* presenting questions both in written text and in spoken forms, and in *NL247* using interactive exercises that involve situated visual aids, such as realistic-looking agendas and letters.

The situatedness requirement (**R3**) represents a special case. All software packages studied use a certain degree of situated content: exercises and examples are embedded in the larger goal of attempting to integrate in the Netherlands. Narrative scenarios are supported with avatar personas, designed to embody and represent the user demographics. This level of situated content represents a 'partial' form of situatedness. The material used is closer to the daily-life experiences of the users than entirely non-situated material, and thus provides the described cognitive, affective, and social learning benefits to some degree. However, it is not fully personalized and contextualized to the individual users. Educational material and content that uses real, immediately recognizable elements from the user's day-to-day life would represent a higher level of situatedness. Future studies should investigate whether or not this level of individual situatedness is practical to achieve, and if the benefits derived from doing so outweigh the additional required effort.

While learning support (**R6**) is seen as important and necessary in literature, practical software implementations are uncommon. As most of the software packages have not been designed as stand-alone learning methods, the learning support is assumed to come from teachers and peers, not from the program itself. While these kinds of learning support are still beneficial, digital learning support has potential benefits of its own. Only the *IJsbreker Plus* and *Naar Nederland* software packages explicitly offer audio support and speech recognition in different languages. *NL247* possesses both a technical support help desk and an easily-accessible dictionary, but limits its content-level automated support to a text message functionality between learners and teachers. Uniquely, *Thuis in Nederlands* offers 'e-coaching' as a method of digital direct-contact support. Given the beneficial claims associated with digital learning support, the relative rarity of software-based support should be investigated in more depth.

Learning style and preference adaptability (**R1**) has proven to be a difficult concept to find and operationalize in practice. *IJsbreker Plus*, *Naar Nederland*, *Thuis in Nederlands*, and *NL247* offer adaptability personalization in the form of a range of different exercise types. The fact that other software packages mostly stick to multiple choice questions suggest that any adaptability in these methods is found in the other materials. Future study could investigate if software-based adaptability has significant benefits over this existing method- and material-based adaptability, particularly where low-literate learners are concerned.

Implementations of social interaction and collaboration (**R4**) functionality are rare. Only *Thuis in Nederlands* offers e-learning functionalities and online group-based exercises. Again, many software packages primarily offer self-study exercises to supplement existing classroom work. The lack of social cooperation options built into the software packages limits their effectiveness as standalone solutions. However, more study is needed to determine whether or not mediated collaboration efforts would work well with low-literate learners to begin with.

Finally, gaming principles (**R8**) are equally rare, only showing up once: *Thuis in Nederlands* uses virtual board games and a virtual reality environment to enhance its learning experiences and engage learners. These findings represent a major departure from literature assumptions, which warrants future study. Schouten, Pfab, Cremers, van Dijk, & Neerincx [150] have already demonstrated that literature expectations regarding gaming principles and gamification might not translate to a low-literate user group.

2.5.3. Enabling Technology

This section describes how VLE technology can serve as an 'enabling technology' for VESSEL. As described in section 2.1, VLEs can contain computer-generated environments and spaces, present digital actors and characters, and possess digital objects and artifacts [38, 39]. These attributes each provide potential benefits to fulfilling

	EHBN	ETV.nl	Naar Nederland	Thuis in Nederlands	IJsbreker Plus	NL247
R1. Adaptability			+	+	+	+
R2. Sensitivity	+	+	+	+	+	+
R3. Situatedness	+	+	+	+	+	+
R4. Collaboration				+		
R5. Multimodality	+	+	+	+	+	+
R6. Support			+	+	+	+
R7. Interactivity	+	+	+	+	+	+
R8. Gaming Principles				+		

Table 2.2: Mapping of the eight requirements on current societal participation learning support software packages. For each software package, each row shows adherence to the listed requirement. The symbol '+' means the software package clearly implements the listed requirement.

section 2.4's functional requirements. Not all current VLEs use all possibilities in equal measure. For example, the *Thuis In Nederlands* Virtual Neighbourhood uses a virtual space and virtual characters, but does not use interactive objects. And some virtual coaching programs [49, 53] focus solely on convincing virtual characters. For each of these three attributes, a description is given of which requirements are more easily implemented by using this technology, and why this is the case. Table 2.3, at the end of the section, also shows a schematic overview of this.

Environments constitute the digital 'spaces' of a VLE. These environments can be abstract, or they can be realistic depictions of existing spaces. A well-designed virtual environment is almost a prerequisite for the success of VLE-based learning [151]. The following six requirements are enabled by this functionality:

- *R1. Adaptability*: VLE environments can provide different spaces for different kinds and levels of exercises. According to Barak [57], this adaptability is much easier to realize in ICT environments than in classrooms: Digital spaces can be altered much more easily than real ones.
- *R3. Situatedness*: Maybe the most intuitive benefit of VLE technology is the potential to deliver a level of spatial situatedness no other software is capable of reaching. Realistic task environments beneficially influence feelings of physical and social presence, situatedness, and learning transfer [39]. Strongly situated virtual environments could be especially useful for participation skills training, the application of which is lacking in current software (see Appendix A). In the Netherlands in particular, this would address a significant gap in the current participation learning curriculum [2, 3, 27].
- *R4. Collaboration*: Virtual environments can provide a shared space for learners to collaborate in, that is time- and location-independent. The *Thuis in Nederlands* software package is an example of this. While R. D. Johnson et al. [105] warn against the socially isolating nature of digital learning, the

high social presence associated with shared virtual environments can actually facilitate the formation of peer connections.

- *R5. Multimodality*: Virtual environments are almost inherently multimodal [83], combining visual information with text and audio.
- *R7. Interactivity*: VLE-based learning set in realistic virtual spaces benefits from intuitive interaction possibilities [83]. Virtual spaces designed to afford realistic interaction are easy to parse even for learners with relatively little computer experience.
- *R8. Gaming principles*: The natural interaction style and increased social interaction offered by virtual spaces [39] all enable the immersion, engagement, motivation, and attitude benefits claimed from the use of gamification in learning [59, 152].

Actors constitute the digital 'characters' of a VLE. Actors can either be *avatars*, serving as digital stand-ins for users, or *agents*, which are autonomous software programs. The following seven requirements are enabled by this functionality:

- *R1. Adaptability*: Agents can adapt their looks and behaviour to better match user wishes and needs. Again, this is easier in a VLE than it is in a real classroom [22, 57].
- *R3. Situatedness*: Actors of any kind, either user avatars or embodied conversational agents, can add situatedness to exercises [153]. These actors can act as conversation partners, or serve as 'social background dressing', adding a layer of affective stress to social situations.
- *R4. Collaboration*: Virtual avatars can allow learners to 'see' and interact with each other remotely. Studies show that the increased social presence that results from learning with other humans, even digitally-represented ones [83] is beneficial to learning effectiveness [154].
- *R5. Multimodality*: Virtual characters naturally present spoken dialogue.
- *R6. Support*: VLE actors can serve as representations of teachers and peers, allowing users to ask for help in a natural way. VLE actors can also serve as autonomous digital characters, offering structured, individualized, computer-guided support. The benefits of this learning support on success and persistence [55, 105], motivation and contentment [155], and self-efficacy [109] are well-documented.
- *R7. Interactivity*: Virtual characters enable a range of scenario-specific affordances for interactivity in learning.
- *R8. Gaming principles*: Cornelissen et al. [59] identify avatar personification and social comparison, two elements commonly associated with virtual characters, as instruments usable for tapping into flow and fiero in a gaming setting.

Objects constitute the digital 'things' of a VLE. The following five requirements are enabled by this functionality:

- *R1. Adaptability*: The presence or absence of VLE objects can change the nature and difficulty of an exercise. An exercise can be made more complicated by the inclusion of hard-to-understand objects, or made simpler by the inclusion of objects that are easy to parse and use.
- *R3. Situatedness*: In any given exercise environment, providing appropriate and realistic tools and other objects can improve situatedness. In exercises where 'learning to handle the object' is the goal (e.g. learning online banking), realistic objects are almost a necessity for success.
- *R5. Multimodality*: VLE objects can present information in many ways, depending on the nature of the object: examples include digital books and newspapers for written information, or digital billboards or televisions for visual information.
- *R7. Interactivity*: Realistic digital objects strongly afford interactivity. ICT objects in particular can be recreated feature-perfectly in a VLE. This allows for incredibly applied practical learning.
- *R8. Gaming principles*: VLE objects like trophies and badges can add a layer of tangibility to gamification-related rewards.

	Cognitive Claims	Affective Claims	Social Claims	VLE Environments	VLE Actors	VLE Objects
R1. Adaptability	+	+		+	+	+
R2. Sensitivity		+	+			
R3. Situatedness	+	+	+	+	+	+
R4. Collaboration	+		+	+	+	
R5. Multimodality	+	+		+	+	+
R6. Support	+	+	+		+	
R7. Interactivity	+		+	+	+	+
R8. Gaming Principles		+	+	+	+	+

Table 2.3: Overview of **(a)** claimed benefits per design requirement, and **(b)** requirement implementation benefits afforded by VLE attributes. For columns 'Cognitive Claims', 'Affective Claims', and 'Social Claims', symbol '+' means that this requirement has claims of benefit associated with this category. For columns 'VLE Environments', 'VLE Actors', and 'VLE Objects', symbol '+' means that this requirements benefits from this kind of VLE implementation.

2.6. Conclusion

This study has used the Socio-Cognitive Engineering method to create a design specification for VESSEL, a virtual environment to support the societal participation education of low-literates. Analyses of the operational demands and human factors provided the foundation of the design: Demographic information, societal participation models, learning processes, theories of adult learning, and computer-supported learning principles. From this foundation, a baseline of eight functional requirements and eighteen associated claims of cognitive, affective, and social learning benefit was derived. Current technology was then assessed using these requirements, in order to highlight discrepancies between literature-backed theory and the best practices of existing societal participation learning support programs. Finally, the advantages offered by VLE technology were described in terms of environments, actors, and objects.

In line with the four main research questions, this paper presents four major results. Research question one was: "*Which attributes of the actor demographics, societal participation behaviours, learning processes, adult learning theories, and computer-supported learning principles, are relevant for the design of VESSEL?*" To answer this question, an overview was created of the problem area of insufficient societal participation for people of low literacy. Demographic information, practice-backed learner profiles, models of societal participation behaviour, and a description of the cognitive, affective, and social processes underlying learning were used to show the operational demands associated with designing in this field.

Research question two was: "*How can these attributes be used to adapt learning support to the skills and characteristics of individual users? Which functional requirements and claims can be derived?*" To answer this question, the operational demands overview and a human factors framework of adult learning theory and computer-supported learning principles were translated into a list of functional requirements for societal participation learning support software. These requirements, justified with theory-backed claims of cognitive, affective, and social benefits to learning, form the specification for the VESSEL system. Table 2.3 shows this specification.

Research question three was: "*Which of these functional requirements are met by current learning support programs? And which ones are not? And what lessons can be learned from this?*" To answer this question, six learning support software packages, taken from Dutch integration learning programs, were assessed on the basis of these requirements. Requirements 'R2. Sensitivity', 'R3. Situatedness', 'R5. Multimodality', and R7. Interactivity' were found in all software packages; examples were presented of best-practice ways of implementation for these requirements. Requirements 'R1. Adaptability', 'R4. Collaboration', 'R6. Support', and 'R8. Gaming Principles' were found sparingly. This finding represents a discrepancy between theory and practice; this was highlighted in terms of the claims associated with these requirements.

Finally, research question four was: "*How can these requirements and lessons be used in the design of VESSEL? And how can the use of virtual environment technology help with this?*" To answer this question, an overview was created of the requirement implementation benefits of VLE technology. Table 2.3 shows that the use of VLE technology has significant benefits over a non-VLE system: VLE environments, actors, and objects enable many requirements to be implemented effectively. It can be concluded that VLE technology is a good technological basis for the proposed VESSEL system.

In closing, this paper offers several clear directions for future study. Results from the requirements assessment show that the requirements 'R1. Adaptability', 'R4. Collaboration', 'R6. Support', and 'R8. Gaming Principles' represent areas of particular interest. While theory and literature show potential benefits in the application of these principles for VESSEL, software in current practice tells a different story. Current societal participation learning support programs do meet the requirements 'R2. Sensitivity', 'R3. Situatedness', 'R5. Multimodality', and 'R7. Interactivity' well, and offer practical examples of how to implement these in the design of VESSEL. Follow-up studies in this field should focus on verifying the practical effectiveness of the proposed VESSEL specification: the translation of existing requirement implementations into a virtual environment solution should be prototyped and tested, and the theory-practice discrepancy regarding learning support, collaboration, adaptability, and gamification should be investigated.

3

Empirical Refinement

Specialized learning support software can address the low societal participation of low-literate Dutch citizens. We use the Socio-Cognitive Engineering method to iteratively create a design specification for the envisioned system VESSEL: a Virtual Environment to Support the Societal participation Education of Low-literates. An initial high-level specification for this system is refined by incorporating the societal participation experiences of low-literate citizens into the design. In two series of user studies, the participant workshop and cultural probe methods were used with 23 low-literate participants. The Grounded Theory method was used to process the rich user data from these studies into the Societal Participation Experience of Low-Literates (SPELL) model. Using this experience model, the existing VESSEL specification was refined: requirements were empirically situated in the daily practice of low-literate societal participation, and new claims were written to explicate the learning effectiveness of the proposed VESSEL system. In conclusion, this study provides a comprehensive, theoretically and empirically grounded set of requirements and claims for the proposed VESSEL system, as well as the underlying SPELL model, which captures the societal participation experiences of low-literates citizens. The research methods used in this study are shown to be effective for requirements engineering with low-literate users.

The content in this chapter has been published as: Schouten, D. G. M., Paulissen, R. T., Hanekamp, M., Groot, A., Neerincx, M. A., & Cremers, A. H. M. (2017). Low-literates' support needs for societal participation learning: Empirical grounding of theory- and model-based design. Cognitive Systems Research **45** (2017) [156].

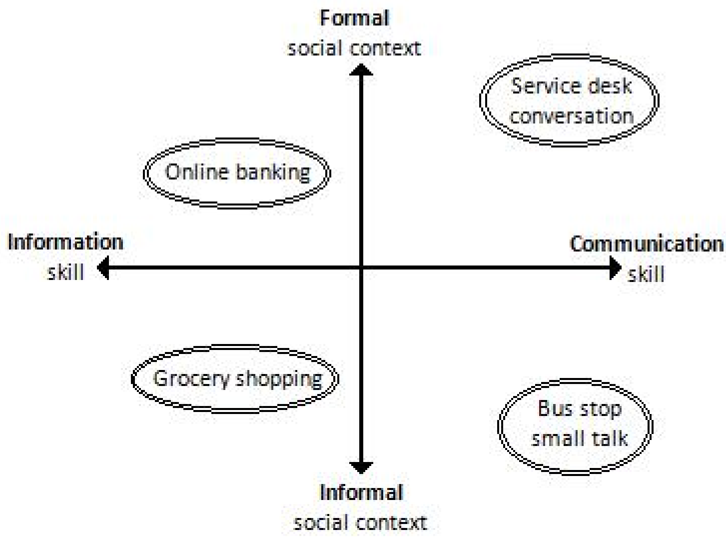


Figure 3.1: Two-axis model of societal participation behavior. Societal participation behavior requires information skill and communication skill, and takes place in formal and informal social contexts. Provided examples in each quadrant represent behaviors that low-literate citizens struggle with. Image from [78]. Behavior examples drawn from [2, 3, 112]

3.1. Introduction

Low-literate citizens of the Netherlands participate in society relatively little [1, 10, 11]. 1.3 million people aged 16-65 (10% of the Dutch labor force, cf. [9]), both native citizens (65%) and non-native citizens (35%), are classified as low-literate; many people in this group struggle with participating, resulting in isolation, unemployment, and low socio-economic status [80, 81]. Societal participation means acting in society to reach certain goals [1]. In the Netherlands, the term 'crucial practical situations' describes those behaviors that are seen as vital for participating independently [2, 3]. Particular *skills* ('core skills of participating in an information society', cf. [4]) are the basis for these behaviors. Reading, writing, speaking, and listening are core skills for getting access to *information*, and for *communicating* with others. These information and communication skills pose different challenges in different *social contexts*. *Formal* social contexts are characterized by rigid rules and impersonal atmosphere, while *informal* social contexts are characterized by flexible social structures and a lack of formal rules [78]. Fig. 3.1 shows a schematic model of the skill and social context dimensions, and includes example crucial practical situations that low-literates may struggle with.

The issues that low-literates can encounter in these situations can be partially explained by a lack of reading, writing, speaking, and listening skills, which are mainly cognitive in nature. Other issues that play a role in societal participation tasks have affective and/or social origins instead [78, 157]. Referring to Fig. 3.1, online banking is an example of a cognitively challenging situation, since a complex website demands high information skill levels from the user. A service desk conversation is affectively challenging, as stress is provoked by an intimidating formal situation and fears about 'making mistakes' and 'not knowing what to say'. Chatting at a bus stop is predominantly socially challenging, since it highlights the effects of limited vocabulary and lack of knowledge about social mores and norms. In practice, all difficult scenarios present a combination of cognitive, affective, and social challenges. Correspondingly, the issues that low-literates face are often multifaceted and highly context-dependent. Different low-literate people will encounter different problems, depending on their cultural background, educational history, vocabulary, and pre-existing skills.

Because of these issues, it is challenging to provide *effective* learning. Traditionally, 'high learning effectiveness' is equated with learning that focuses on desired, meaningful learning outcomes [34, 158–160]. However, this overlooks both people that drop out of the learning process partway, and people that never start learning at all. De Greef, Segers, Nijhuis & Lam [10] show that for low-literate learners, starting and persisting with learning are non-trivial challenges. Based on student learning models by Biggs & Moore [161] and Cybinski & Selvanathan [162], we use a broad view of learning effectiveness: Learning is effective if learning *accessibility* is such that learners have no significant barriers to start [34–36], if the learning *experience* matches learner skills, needs, and wishes [36, 159, 162], and if learning *outcomes* show that meaningful, desired learning goals are reached [34, 158–160].

In the Netherlands, the available learning options for low-literates can focus alternatively on language learning (vocabulary and grammar), societal knowledge (rules and norms of Dutch society), and participation skills (the information and communication skills needed to participate in formal and informal contexts), as needed by the individual learner [78]. These learning options aim to provide the language, rules, and knowledge of practical skills needed to handle crucial practical situations. This learning mainly takes place in classrooms. De Greef [32] shows that classroom lessons increase social inclusion and societal participation for those who complete them, showing that this method of learning helps learners reach desired learning outcomes and suggesting that the learning experience is positive enough to keep learners engaged. However, class-based methods can be difficult for some low-literates to access: barriers might be physical (learners have to get to the classroom), emotional (learners can have bad experiences with 'going back to school'), social (learning might not work well in a group of people), or financial (learners might not be able to afford tuition). Correspondingly, we aim to design accessible self-learning support that can complement existing classroom practices.

Furthermore, we design this support to address the different learning needs of the heterogeneous 'low-literate' demographic. Commonly, a division is made between 'first-language' learners, native speakers primarily struggling with reading and writing, and 'second-language' learners, who are non-native citizens whose issues pertain to vocabulary and listening to spoken language of their non-mother tongue. In the Netherlands, these groups are referred to as 'NT1' (meaning 'Dutch as a first language') and 'NT2' (meaning 'Dutch as a second language', cf. [80, 81]. Kurvers, Dalderop, & Stockmann [89] provide a more detailed classification of five ‘learners profiles’, which builds off the NT1/NT2 division and adds background attributes to define five meaningful subgroups (we describe this in more detail in [78]). Different groups encounter different problems in different situations: for example, NT1 learners commonly have no problems with speaking and listening to Dutch, but struggle with ICT more than NT2 learners. It is difficult for class-based methods to personalize to particular demographics or to individual learners, or to use content that is equally difficult to all groups. This problem is exacerbated if learners can join or leave classes at will [10]. A learning method that allows for continuous and comprehensive personalization could be a valuable addition to current classroom practices.

We believe that learning support software can provide this addition. Learning software is already used by many low-literate classes (see [78] for an overview), where the software packages tend to be an integrated part of the classroom lesson plan. As a complement, we aim to design individualized, situated learning support software. We call this envisioned system ‘VESSEL’: a Virtual Environment to Support the Societal participation Education of Low-literates. Realistic scenario-based learning (cf. [22]), using content drawn from real daily life experiences, can be incorporated into VESSEL. This can allow individual low-literate learners to practice skills and behaviors that are relevant to them, and (by doing so) acquire language skills and societal knowledge in the proper context of use, wherever and whenever they want. The current study focuses on the generation of an empirically grounded design specification for VESSEL. For this, we need a comprehensive understanding of the context of low-literate societal participation learning: The societal participation domain, the demographic of low-literate citizens, the issues, and the potential solutions. This understanding must then be translated into a specification that describes the system’s intended functionality and design rationale. We use the Socio-Cognitive Engineering (SCE) method for this [84, 85], an iterative software design and development method consisting of a *foundation* phase wherein relevant data are collected, a *specification* phase wherein a design specification is created, and an *evaluation* phase wherein the design specification is evaluated [163]. In previous work, we collected a foundation of relevant knowledge based on literature and theory, and created a high-level VESSEL design specification (shown in Table 3.2, cf. [78]). However, this specification does not yet address the subjective lived experiences and problems of low-literate citizens, such as accessibility issues low-literates encounter with regard to learning, examples and behaviors they want to see reflected in the learning experience, and skills and experiences they see as desirable learning outcomes. Incorporating qualitative, empirical user data into the operational demands provides this insight, and better situates the foundation in the problems and experiences of low-literates. This lets us *refine* the existing specification (cf. [73]) to more accurately reflect low-literates’ needs and desires about societal participation learning, grounding it both in theory and in the empirical practice of daily life.

SCE can be considered as a repository of findings that are collected and described systematically over the course of a software development trajectory. Depending on the research context, user demographic, and application domain, appropriate methods can be chosen to gather foundation data, create specification requirements, and evaluate claims. We selected our methods to meet three goals. First, we want to evaluate our current theoretical insights about low-literate societal participation, to see if our assumptions and ideas are reflected in current practice. Second, we want to see the real daily life experiences of low-literates, and learn about the practical ways low-literates approach societal participation. Finally, we want to use these theoretical evaluations and new insights to refine our initial VESSEL specification. Our data acquisition methods must take the skills, needs, and limitations of the target group into account [71, 82, 150], and our data analysis methods must be suitable for working with rich empirical data. Three methods have been selected: the data acquisition methods 'participant workshop' [71] and 'cultural probe' [72], and the data analysis method 'Grounded Theory' [74]. We think these methods are suitable for gaining insight into the daily lives of low-literates, and refining our VESSEL specification. However, to the best of our knowledge, there have been no practical applications of these methods with a low-literate target group for the purposes of requirements engineering. Consequently, two research questions emerge:

1. How can the research methods 'participant workshop', 'cultural probe', and 'Grounded Theory' be used to map out the subjective societal participation experiences of low-literate citizens, and to update the operational demands with qualitative empirical data?

2. Which VESSEL specification refinements are found by applying these methods?

In the proceeding, we answer these questions in the following way. We first provide background on the SCE methodology, including a detailed description of relevant design specification concepts and an overview of the design specification drawn from our previous work, and an introduction of our chosen methods 'participant workshop', 'cultural probe', and 'Grounded Theory' (section 3.2). We then explain how the methods were adapted to and used with our low-literate target audience to collect and analyze empirical user data (section 3.3). We show how these data were used to create the resulting 'Societal Participation Experience of Low-Literates' model (SPELL) and to refine the existing VESSEL design specification (section 3.4). Finally, we present conclusions, limitations of the current study, and directions for future work (section 3.5).

3.2. Background

3.2.1. Socio-Cognitive Engineering

The SCE method (Fig. 3.2) is an iterative software design and evaluation method that moves non-linearly through three phases. In the foundation phase, relevant data are collected: This includes *operational demands* (descriptions of user demographics, application domain, and system tasks), *human factors knowledge* (relevant theory about user-system interactions), and *technology* (both existing and

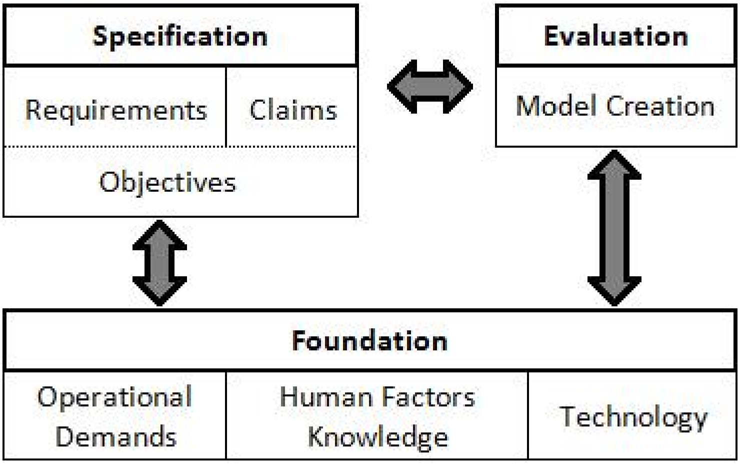


Figure 3.2: Graphical representation of the Socio-Cognitive Engineering method. Method from [84, 85]

envisioned technology). In the specification phase, these data are used to create a design specification. As part of this specification, *objectives* describe the general operational or domain goals of the software system that is designed. *Requirements* describe the system's intended functionality, i.e. 'what the system should do'. Each requirement is associated with one or more objectives. *Claims* explicate the expected operational outcomes of implementing each requirement (positive or negative) into variables. Requirements describe what the software should do in order to reach system objectives; claims describe why [84, 85, 163]. In the evaluation phase, the claims are evaluated in literature study or experiments with end users. In Schouten et al. [78], we used the SCE method to gather a foundation of data (a description of the low-literate target demographic in the Netherlands and of VESSEL's context of use, an overview of existing learning support software, and relevant adult learning and computer-supported learning theory, cf. [78]) and create a high-level design specification for our envisioned VESSEL system, consisting of eight functional requirements and 25 claims. Table 3.2 shows short descriptions of each requirement.

The main objective of VESSEL is to offer effective societal participation learning to low-literate learners. This high-level objective can be decomposed along the two dimensions described in section 3.1: The 'broad learning effectiveness' dimension of accessibility, experience, and outcomes, and the 'societal participation' dimension of cognitive, affective, and social aspects. This results in a three-by-three overview of nine sub-objectives, shown in Table 3.1.

Each column in Table 3.1 shows three objectives, cognitive and affective and social, related to one dimension of learning effectiveness. The first column of three objectives relates to the accessibility of the offered learning. Learning accessibility is particularly important for low-literates, since (as adults) schooling is not compulsory for them. Learners must be able, ready, and willing to start learning. **O1.**

	Accessibility	Experience	Outcomes
Cognitive	O1. Difficulty	O4. Performance	O7. Success
Affective	O2. Barriers	O5. Valence	O8. Self-efficacy
Social	O3. Intention	O6. Engagement	O9. Retention

Table 3.1: Nine sub-objectives, derived from the main objective of 'societal participation learning effectiveness for low-literates'. Left to right: Objectives apply to either learning accessibility, learning experience, or learning outcomes. Top to bottom: Objectives apply to either cognitive aspects of learning, affective aspects of learning, or social aspects of learning.

Difficulty means that the learner is cognitively able to access and use the learning: the learning is not too difficult, and does not require skills the learner does not have. **O2. Barriers** means that the learner is emotionally ready and able to engage with the learning: emotional barriers like fear and shame have been alleviated or removed. **O3. Intention** means that the learner is socially ready to use the learning: The learner has the intention and the motivation to start learning.

The second column of three objectives relates to the experience of doing the offered learning. This experience must be tailored to the needs, wishes, and issues of low-literate learners. Learners must be able to comprehend the learning, but they must also feel positive and motivated. **O4. Performance** means that the learner can cognitively follow and interact with the learning: The learning matches the learner's skills and abilities. **O5. Valence** means that the learner is emotionally able to keep up with the learning: The learning has a positive valence, and does not place strong emotional demands on the learner. **O6. Engagement** means that the learner is socially prepared to follow the learning: The learner is engaged and motivated to play a part in the process.

The third column of three objectives relates to the outcomes of the offered learning. These outcomes must go beyond 'only' test scores: Stimulating low-literates to participate more requires not only skills, knowledge, and vocabulary, but also a positive self-image and a willingness to continue learning throughout [78]. **O7. Success** means that the learner reaches measurable learning goals: The learner learns and recalls content, and/or gains the skills and knowledge intended to be transferred. **O8. Self-efficacy** means that the learner increases their self-efficacy with regard to societal participation: They become more confident about their ability to successfully complete crucial practical situations (cf. [109]). **O9. Retention** means that the learner is more motivated to return to this learning later, or find other learning themselves.

Table 3.2 shows the high-level VESSEL specification from Schouten et al. [78]. The eight functional requirements are shown. The table also shows which of the nine Table 3.1 sub-objectives each requirement contributes to. Each connection between one requirement and one objective makes up one high-level claim: The *claim* is that meeting this requirement will beneficially contribute to reaching this objective. For example: On the second row, requirement 'R1. Adaptability' is associated with two objectives: 'O4. Performance' and 'O5. Valence'. This means

3

Requirement	Objectives
R1. Adaptability: VESSEL should offer and/or support different learning styles and preferences.	O4. O5.
R2. Sensitivity: VESSEL should use non-confrontational language language and content, demonstrate cultural awareness, and take existing emotional issues with regard to reading and writing and societal participation into account.	O5. O9.
R3. Situatedness: VESSEL should use learning materials and contents that are closely related to the learner's physical environment and real-life experiences.	O2. O7. O9.
R4. Collaboration: VESSEL should have systems in place that enable, support, and foster social interaction and collaboration in learning.	O2. O3. O7. O9
R5. Multimodality: VESSEL should employ multimodality, offering content in multiple concurrent ways.	O1. O2. O4.
R6. Support: VESSEL should possess built-in support options.	O4. O6. O7. O8. O9.
R7. Interactivity: VESSEL should employ real interactivity in offering content.	O4. O6. O7.
R8. Gaming principles: VESSEL should use elements and principles of interactive gaming.	O5. O6. O9.

Table 3.2: VESSEL specification, from [78]. Left column: requirement name and description. Right column: the objectives this requirement is claimed to influence.

that 'R1. Adaptability' has two associated claims: A VESSEL design that meets this requirement will lead to *better performance*, and *more positive valence*. Requirement 'R2. Sensitivity' shares the same claim 'O5. Valence'. This means that both requirements are *independently* predicted to result in a more positive valence.

3.2.2. Justification / Adaptation of Data Acquisition and Analysis Methods

Participant Workshops

Participant workshops combine the semi-structured group discussion style of focus groups (cf. [71]) with structured group exercises to acquire data. This combination of techniques provides rich data: The discussion segments allow participants to share their own experiences and insights, while the exercises ensure that certain questions are answered in a structured way. A workshop leader supervises the discussions and the exercises, ensuring that all group members get a chance to contribute and solving problems as they arise. The proceedings are also recorded, for later analysis.

Participant workshops were selected for use in this study because they let us cover pre-selected topics of interest. Furthermore, workshops can cover the whole spectrum of the societal participation experience: This includes cognitive topics, but also affective and social ones. A combination of directed questions, structured exercises, and group discussions reveals all relevant aspects of societal participation. Finally, participant workshops seem to match the limitations of low-literate participants well. Curated group discussions allow all participants to provide input without requiring any reading and writing skill, and using a spoken language complexity level that matches participants' current skills. Any exercises used will have to take the reading, writing, speaking, and listening limitations of the target group into account: Possible examples are visual exercises using graphical elements, or 'think-aloud' brainstorm exercises using non-complex language (cf. [15]).

Cultural Probes

The cultural probe method is a qualitative data acquisition method, focused on collecting unstructured, subjective, individual experiences [72]. In a cultural probe study, the responsibility for data collection is placed with the participants. Participants receive a 'cultural probe kit' at the start of the study, which contains tools and methods for recording data: Common examples include maps, diaries, voice recorders, and disposable cameras [71, 164, 165]. The probe kit also contains instructions on how and when to use the probe. Instructions range from highly detailed to open-ended: While Schmehl, Deutsch, Schrammel, Paletta & Tscheligi [166] have used directed cultural probes to answer specific research questions, Gaver et al. [164] argue that the cultural probe's real value lies in providing complex, uncertain, hard-to-interpret data.

Cultural probes were selected for use in this study because they let us take unsupervised looks into the real daily lives of participants. This method in particular seems useful for engaging low-literate participants in the research process. Traditional text- and detail-heavy study methods (*e.g. questionnaires*) can run into barriers of information complexity [150], negative participant attitudes [91], and low engagement and motivation. However, a probe kit can be designed to minimize the need for reading and writing skill in favor of more 'playful', engaging activities [164]. By using other types of materials, such as photos, recordings, and drawings, it is possible to overcome cognitive, affective, and social barriers to using the probe. This makes it more likely that participants will actually use the method in their daily lives.

Grounded Theory

Grounded Theory is a qualitative data analysis method aimed at theory generation, built on the idea that theories need to be 'grounded in data'. Data drives the whole research process, and theories are derived directly from data [74, 167–169]. Although the Grounded Theory method is mutable and open to interpretation, a set of basic elements can be distinguished (cf. [168, 170–172]). Grounded Theory is commonly a part of an iterative process of data collection and analysis. New data is transcribed and coded on a conceptual level. Initially, an open coding procedure [172] is used: The aim is to identify as many interesting concepts as possible. New data sets and codes are constantly compared to older ones, a process known as the constant comparative method [167, 168, 173]. Codes are collected in categories and higher-order categories. This gradually creates a hierarchical overview of the data, and helps to identify the study's core variable: The particular code or concept that accounts for most of the variance related to the area of study [172]. Once identified, the core variable becomes the focal point of later work, and coding switches to selective coding [172]: The aim is to link all codes and categories to the core variable, imposing a hierarchical ordering on the data. Researchers write memos throughout the study: code notes describe discovered concepts in as much detail as possible, and theoretical notes hypothesize about the relationships between found codes and categories (cf. [171]). These memos support the incorporation of new findings into existing literature [167].

Grounded Theory was chosen for this study as it seems suitable for transforming societal participation user experience data into functional requirements and claims. Grounded Theory is well-suited for dealing with rich data [167]. The conceptual modeling techniques that are part of this method [171] impose a formal hierarchical ordering on varied and expansive data. Additionally, Grounded Theory results meet Lincoln & Guba [174]'s four determinants for qualitative data correctness: confirmability ('*are study results and claims really reflected in the data*'), credibility ('*are study results and claims factually correct*'), transferability ('*do study results and claims show up in the real world*'), and dependability ('*can study claims and results be repeated by other researchers*'). Grounded Theory results are *confirmable* because they are completely 'grounded' in their data. They are *credible* because all data reflect things that participants have directly said and done. And they are *transferable* and *dependable* because repeated data collection and theoretical sampling (running many data collection sessions as the need arises) leads to a saturated, as-complete-as-possible overview of the problem space. Consequently, Grounded Theory results are rich, reproducible, and accurate about the state of the world.

In this study, however, a shorter version of the Grounded Theory method is used. Because of practical limitations, the element of theoretical sampling was scrapped entirely: Instead of planning data acquisition sessions on a per-need basis, a full plan of sessions was drafted in advance and carried out as-designed. The constant comparative method was only used once, after final completion of the data acquisition. This makes our method is different from the traditional format seen in literature [74, 167, 170–172, 175]. Again using Lincoln & Guba [174]'s determinants, the expected effects of this choice can be traced. Our shorter method performs as well as longer Grounded Theory on the confirmability and credibility determinants: The results from shorter Grounded Theory are similarly 'grounded in data', and directly reflect participants' sayings and actions. However, our method struggles with transferability and dependability. Classic Grounded Theory employs theoretical sampling and running over many iterations to gain a fully saturated overview of the problem space. This is not reflected in our shorter method.

Steps can be taken to alleviate these problems. Insufficient dependability indicates that researcher bias may have a significant effect. This can be minimized by involving multiple researchers in all steps of the process. Insufficient transferability indicates that the study results might not 'show up in the real world'. The strongly heterogeneous nature of the low-literate demographic (cf. [78]) makes this an important issue. The impact of this can be minimized by carefully designing studies and selecting participants to cover as much variety in the target demographic as possible. We use these two measures to try and ensure qualitative data correctness in our shorter Grounded Theory method; we reflect on the effectiveness of this in section 3.5.

3.3. Methods

3.3.1. Data Acquisition

To collect sufficient qualitative data about low-literate societal participation, the participant workshop and cultural probe methods were combined into a structured study series. This study series was carried out twice, with two different groups of participants, as 'Study One' and 'Study Two'. Participants were selected to match profiles 2, 3, and 4 (ranging from low-medium to medium-high skill levels) of Kurvers, Dalderop & Stockmann [89]'s five language learner profiles, which can be used to divide the low-literate demographic according to (cultural) background, language level, and educational history (by treating all low-literates as first- or second-language learners). Profiles 1 and 5 were not included: Learners that fit profile 1 are too self-sufficient to benefit much from systems like VESSEL, while learners that fit profile 5 are too reading-and-writing-skill-deficient to access and use VESSEL's computer-based learning support. The study series consisted of three elements: One participant workshop focused on societal participation behavior ('societal participation workshop'), one cultural probe study, and one participant workshop focused on technology use ('technology use workshop').

Societal Participation Workshops

Goal. Two societal participation workshops were organized. The main goal was to elicit feedback on particular examples of societal participation behavior; a secondary goal was to introduce and explain the cultural probe (section 3.2.2). Example behaviors were selected in consultancy with experts in the field of low-literate societal participation: these experts included Dutch language teachers, and specially-trained 'literacy ambassadors' (former low-literates with particular training, cf. [77]). Four behaviors were chosen from the list of crucial practical situations (section 3.1) to correspond to the four quadrants of Schouten et al. [78]'s societal participation axis model (Fig. 3.1). This process was repeated in both Study One and Study Two, resulting in two lists of four behaviors. There are two differences between Study One and Study Two: 'reading medicine leaflets' in Study One was replaced with 'reading municipality letters' in Study Two, and 'meeting up with friends' in Study One was replaced with 'talking to neighbours' in Study Two. Table 3.3 shows the selected behaviors.

Participants. In study one, nine low-literate participants (five men, four women) took part. Eight participants were Dutch natives, while one was non-native. Ages ranged from 42 to 62 (M=52.3, SD=6.6). In study two, fourteen low-literate participants (five men, nine women) took part. Ages ranged from 20 to 54 (M=33.5,

	Study One	Study Two
Formal-Information	'Reading medicine leaflets'	'Reading municipality letters'
Formal-Communication	'Calling bank customer service'	'Calling bank customer service'
Informal-Information	'Buying groceries on sale'	'Buying groceries on sale'
Informal-Communication	'Meeting up with friends'	'Talking to neighbours'

Table 3.3: Societal participation behaviors used to elicit group discussion during the Societal Participation Workshops. Each behavior fits one quadrant of the societal participation axis model: formal-information, formal-communication, informal-information, and informal-communication.

SD=9.2). All participants were non-native Dutch citizens. All participants were recruited from language learning classes from the ROC Nijmegen school in the Netherlands, and had similar literacy levels.

Materials. One PowerPoint presentation was used as supporting material. Two voice recorders were used to record the proceedings.

Procedure. Workshops were held in local classrooms, and lasted one hour. People present included: four researchers, the participants, and one 'facilitator', a teacher known to the participants. The workshop started off with a short introductory PowerPoint presentation from the researcher designated as workshop leader, introducing the researchers and the workshop goal to the participants. Participants were shown summary PowerPoint slides of the four societal participation behaviors. After each slide, participants were encouraged to share personal stories about each subject with the group. Participants spoke one at a time. The workshop leader ensured that participants did not start cross-talking, that participants remained broadly on topic, and that individual speaking turns did not run overlong. Other researchers took detailed notes.

Finally, the researchers introduced the cultural probe. The goal of the cultural probe was explained, and the individual elements of the probe kit were demonstrated. In both studies, all present participants agreed to take part in the cultural probe study.

Cultural Probe

Goal. Two cultural probe studies were organized. The goal was to supplement the insights gained from the societal participation workshop with personal, lived experiences, documented by each participant at the moment of occurrence. Participants were specifically asked to look out for two kinds of societal participation occurrences: one example (each day) of something related to societal participation and reading and writing skill that they were *proud* of, and one example of something they had *difficulty* with. This way, both the positive and negative aspects of the societal participation experience were explicitly indicated as important.

Participants. In both studies, all participants who took part in the societal participation workshop agreed to take part in the cultural probe study as well.

Materials. The cultural probe kit was designed in collaboration with the experts (the same language teachers and 'literacy ambassadors', cf. [77]. Each kit consisted of three elements: a disposable camera, a handheld voice recorder, and a set of pre-printed worksheets. These worksheets consisted of nine pages: one page containing a short summary of the cultural probe assignment, two pages serving as short manuals for respectively the camera and the voice recorder, and six pages providing day-specific recording space with two writing areas labeled 'proud' and 'difficult'. Participants could also reference photo numbers in their writing.

Procedure. Each participant was given a cultural probe kit at the end of the societal participation workshop. Participants were informed that the cultural probe study would last for one week, and they were asked to record at least two occurrences every day: one positive occurrence and one negative occurrence. These questions were as open-ended as possible, while still keeping study focus on the experience of participating in society as a low-literate citizen. At the end of the week, participants could turn in the cultural probe kit to the teacher who was present at the workshop. These cultural probe kits were then sent to the researchers. The researchers compiled and analyzed these data in preparation for the following workshop.

Technology Use Workshop

Goal. Two technology use workshops were organized, with two goals. First, the workshops were used to reflect on the cultural probe outcomes. Second, they intended to gain insight into the participants' technology use and to elicit participant feedback on the ICT devices and services that are present or absent in their day-to-day lives. Table 3.4 provides an overview of the specific devices and services that were asked about. Based on feedback after Study One, two new elements were added for Study Two.

Participants. In study one, six of the nine low-literate participants (three men, three women) who completed the societal participation workshop and cultural probe took part. Five participants were Dutch natives, while one was non-native. Ages ranged from 42 to 62 (M=53.3, SD=7.6). In study two, ten of the fourteen low-literate participants (four men, six women) who completed the societal participation workshop and cultural probe took part. Ages ranged from 20 to 54 (M=33.2, SD=9.7). All participants were non-native Dutch citizens.

Materials. One PowerPoint presentation was used as supporting material. This presentation incorporated anonymized pictures and transcribed audio from the cultural probes. Two voice recorders were used to record the proceedings. Additionally, two A2 format posters were used as supporting material for the workshop exercises. Each poster showed two bisecting axes. The horizontal axis measured ease-of-use for devices and services: it was labeled 'easy' on the left end and 'difficult' on the right end. The vertical axis measured degree-of-use for devices and services: it was labeled 'do use' on the top, and 'do not use' on the bottom. Several sets of small stickers, each containing a visual depiction of one device or service, were also used. See Fig. 3.3 for an example of these posters and stickers.

Procedure. Workshops were held in local classrooms, and lasted one hour. People present included: four researchers, the participants, and one 'facilitator', a teacher known to the participants. Each workshop was divided into two parts. During

ICT Devices	ICT Services
TV	Internet
Landline phone	Email
Cellphone	SMS (phone texts)
Smartphone	Skype
Desktop computer	Chatting
Webcam	Facebook
Car navigation system ('TomTom')	Twitter
Tablet computer	Online banking
Laptop computer	Online shopping
DVD player	Computer games
Digital camera (*added in Study Two*)	Using computer for work and learning
	Using computer for route planning
	Photo/video editing
	Customer service contact (*added in Study Two*)

Table 3.4: ICT devices and services presented during the technology use workshop.

the first part, researchers and participants viewed and discussed the results from the cultural probes. Results were viewed anonymously: participants could come forward as the creator of certain material and explain themselves further, but this was not mandatory. All present material creators chose to reveal themselves. The cultural probe examples were used primarily to elicit stories and examples from participants. The workshop leader ensured that participants did not start cross-talking, that participants remained broadly on topic, and that individual speaking turns did not run overlong. Other researchers took detailed notes. The second part of the workshop consisted of a technology use exercise. Participants were asked to provide their preferences on the ICT devices and services listed in Table 3.4. Each participant was given one set of ICT device and service stickers. The two A2-size axis posters were put up in the center of the room. Participants were asked to attach their stickers to each poster, in the quadrant that best reflected their relationship with the ICT device or service depicted on the sticker. One poster was used for all ICT device stickers, and one poster was used for all ICT service stickers.

3.3.2. Data Analysis

The participant workshops and cultural probes resulted in three types of data: audio data, textual data, and graphical data. Workshop proceeding recordings were the principal source of *audio* data. Each of the four participant workshops was recorded from start to finish; resulting audio files varied in length from 45 minutes to 2 hours. Additionally, cultural probe returns included 94 audio files. These files contained recordings of individual participants' experiences, often directly referencing the pictures or writing included in the probe. Probe audio files varied in length from several seconds to about a minute. Cultural probe returns were the

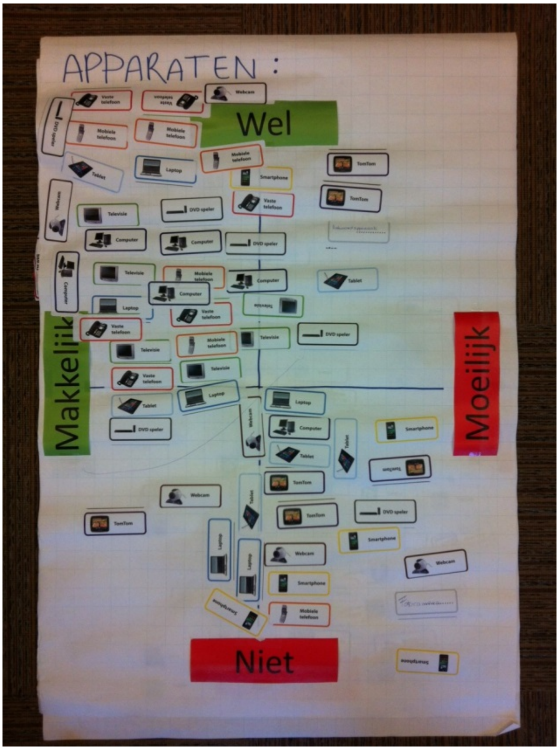


Figure 3.3: The 'ICT devices' poster used in the Technology Use workshop of study one, including participant sticker placement. Text on the poster is in Dutch. Blue text in top left reads 'Devices:' Axis labels read: 'Do' (green top), 'Do Not' (red bottom), 'Easy' (green left), and 'Difficult' (red right)

principal source of *textual* data. 129 text fragments were included in the probes. Text fragments were usually one or two sentences long, though some fragments reached a dozen lines in length. Finally, two sources of *graphical* data were identified. Firstly, cultural probe returns included 157 pictures. Secondly, the technology use workshops resulted in two filled-out 'ICT services' posters, and two filled-out 'ICT devices' posters (Fig. 3.3).

The Grounded Theory methodology was used to transform this data into a hierarchical model of the societal participation experience of low-literate citizens. First, data from the participant workshops and cultural probes was processed for use in the analysis. Audio data from workshop recordings and cultural probe voice recorders was transcribed by three researchers. Written data from notes and cultural probe worksheets was copied to digital text. Graphical data from posters was counted and tabulated. Graphical data from cultural probe cameras was not 'processed', but was used as-is.

Workshop and cultural probe transcripts were analyzed in chronological order by one researcher, using the *open coding* procedure. After all transcripts were coded once, one iteration of the *constant comparative method* was used: A second round of coding was done, using all codes identified so far. This was done to ensure that every transcript was coded from the same final frame of reference. No new codes emerged during the second round of coding.

Because of the focus of this study, it was known from the start that the core variable was 'the societal participation experience of low-literate citizens'. Building on this, the *selective coding* procedure was used to impose a hierarchical ordering of categories on the data. Five researchers engaged in brainstorming sessions to identify this ordering. All conceptual codes were evaluated on accuracy of description, and relevance to the core variable. This ensured quality of fit to the emerging model. Codes that fit together conceptually were combined in new categories. Codes that did not fit other codes, but that were still relevant to the core variable, were made into categories of their own. Codes that emerged to not be as relevant to the core variable as initially thought, and codes where the researchers could not agree on an accurate description, were deleted from the analysis. Code quantity of occurrence played no role here: Codes were accepted or rejected regardless of how often they showed up. The cultural probe camera pictures were used during this process to illustrate and contextualize the data, helping the brainstorming researchers reach agreement.

3.4. Results

3.4.1. SPELL Model

Using the data and analysis from section 3.3, we have created the Societal Participation Experience of Low-Literates (SPELL) model. The SPELL model is made up of four categories that describe the societal participation experience of low-literate Dutch citizens: 'Personal', 'Informal', 'Formal', and 'Interaction'. Each of these first-order categories can contain any number of second-, third-, and fourth-order categories. Each category branch terminates in one or more concepts: concepts form the link between the model and the raw data. Each concept currently in the model (179 in total) is uniquely important and interesting for describing the societal participation experience of low-literate citizens in general, and NT1 and NT2 learners specifically. Fig. 3.4 shows only the first- and second-order categories. In the next section, we explain how the model was created from the data, and describe these first- and second-order categories in more detail. For readability reasons, the complete set of model data (including third- and fourth-order categories and concept descriptions) has not been added to this article. Instead, this model is made freely available as an interactive PDF. This PDF includes expansive descriptions for each category, describes the way each coded concept contributes to the creation of the category, and identifies meaningful links between categories. Photo results from the cultural probes and participant quotes have also been used, to provide context. The PDF is available at: `http://ii.tudelft.nl/sites/default/files/LLSNSPL--Appendix_A.pdf`. Fig. 3.5 shows one example page from this PDF.

Four large clusters were identified during brainstorming: '*Personal*', '*Informal*', '*Formal*', and '*Interaction*'. First, researchers agreed that the largest cluster could be

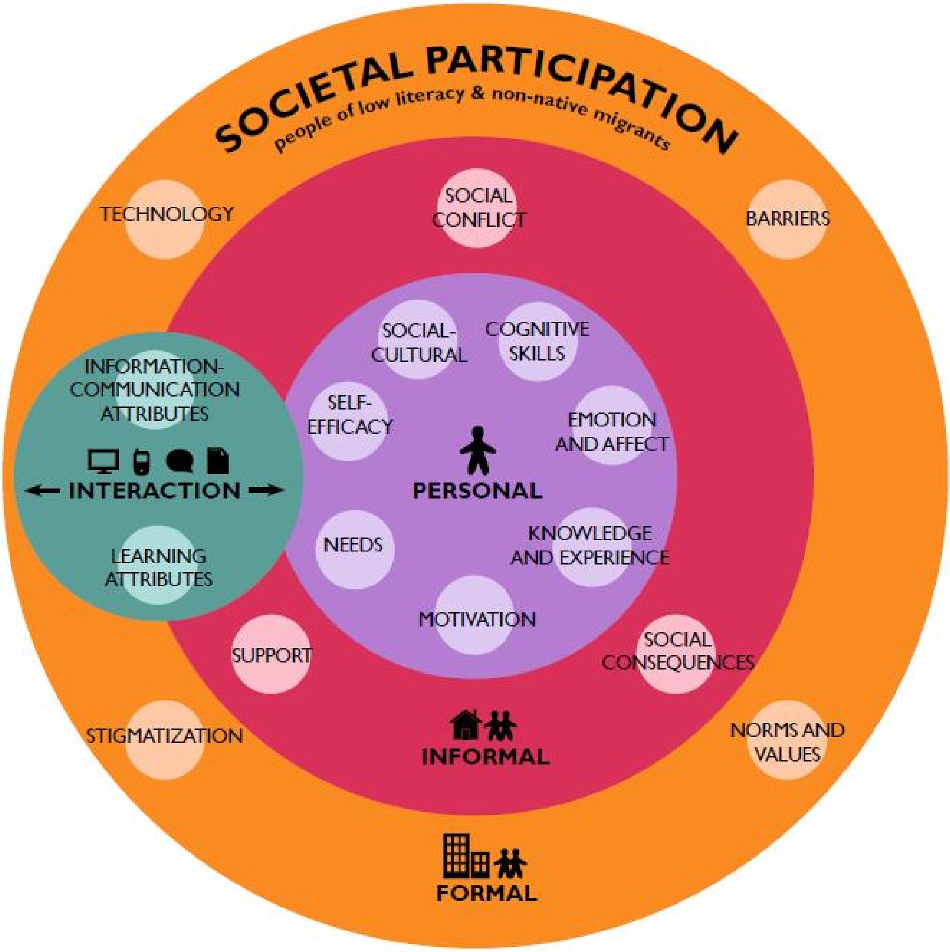


Figure 3.4: The Societal Participation Experience of Low-Literates (SPELL) model

made up of data and codes reflecting on the participants directly. Low-literates very often talked about their own skills, feelings, motivations, educational histories, and experiences with societal participation, addressing cognitive, affective, and social aspects. Participants acknowledged their cognitive difficulties: For instance, one participant said "*Traffic is really difficult for me, in certain situations I can't read traffic signs quickly enough […] it's just incredibly busy, and you just can't read things fast enough.*" An example of an affective problem was seen when one participant described their personal fear of going to school, saying that "*…it's crazy, but I thought… there's going to be someone there […] That's the fear you have.*" Another participant expressed frustration with slow learning process: "*I want to do things much faster, but then, then I can't do it, and then you run into this wall again.*" As an expression of social issues, participants acknowledged that their own limited motivation to 'deal with' low-literacy issues is a major factor ("*If they say that sort of stuff to me […] then I'm done with it as far as I'm concerned. I can live without that*"), but that they do have various needs that are currently not fulfilled ("*I go to school because I want to learn to speak and write better*"). These codes were all collected in the first first-order category, **Personal**. The codes were further clustered into seven second-order categories:

- *Cognitive Skills*: Skills that directly influence the societal participation experience, like language skills, coping strategies, and the math dimension of low literacy.
- *Knowledge And Experience*: Overview of how the presence or absence of societal knowledge, technology knowledge, and language mastery impact participation. This includes (e.g.) the learner's academic and cultural backgrounds, and their operant knowledge about societal participation.
- *Emotion And Affect*: Important emotional reactions associated with low-literate participation: fear, frustration, shame, stress, surprise, and pride.
- *Self-Efficacy*: Different kinds of self-efficacy information that influence the experience of participation, including self-efficacy with regard to understanding information, self-efficacy with regard to technology, self-efficacy about societal participation, and 'general self-efficacy'.
- *Motivation*: Factors that influence the decision to participate on different levels: this includes motivation to learn and to participate, the choice to stop a difficult behaviour or to persist, and 'the threshold', a participant-created code that describes the high affective and social difficulty of going to class for the very first time.
- *Needs*: Those things low-literates need that drive them to participate in society. Examples include the need to learn, the need for social interaction, and the need for attention and validation.
- *Social-Cultural*: Overview of how a low-literate's specific cultural background can influence participation in Dutch society. NT1 low-literates express a sense

of isolation and feeling alone, while NT2 low-literates encounter cultural differences between their country of origin and the Netherlands.

The second largest cluster researchers quickly identified involved codes that reflected on society, specifically the aspects of society that low-literate participants commonly interacted with. The codes were initially treated as a single category, but the researchers could not agree on an accurate description. Using expert input and literature, as well as Schouten et al. [78]'s axis system (see Fig. 3.1), two smaller categories were instead created, reflecting two levels of ideas of 'society' and 'participation'. Based on Fig. 3.1, we called these categories Informal and Formal. The **Informal** category was aimed at day-to-day participation on a smaller scale, which involved interacting with friends and neighbours and participating in the neighborhood and local institutions. Low-literates were more positive about this level of interaction, and talked about benefits of local support: this included functional support ("*I thought this letter was a little difficult because I couldn't understand everything about it, but my sister helped me out and now everything is okay*") and moral support ("*...99% will say that's really positive, good job guy, keep it up! And... and I like hearing that*"). However, social conflicts were often mentioned as well. For NT1 learners, friends and family are not always understanding of low-literacy ("*What I sometimes do is send mails to my father [...] he tells me, you made over twenty mistakes*"); for NT2 learners, neighbours and local contacts are not always willing or able to communicate at the right level of Dutch ("*Yeah, it's difficult for me because in my town [...] they only write in Dutch, for Dutch people*"). The codes were further clustered into three second-order categories:

- *Support*: The positive and negative aspects of being supported by friends, family, and neighbours. Support enables low-literates to survive and to participate, but too much support or poorly provided support can hamper learning and impact social relations.
- *Social Consequences*: Overview of how being low-literate influences participation in daily life. This includes the difficulty of quickly dealing with information, reliance on a (sometimes) small support network, and a pervasive social judgment about low-literacy.
- *Social Conflict*: Overview of how being NT2 low-literate can be a source of informal social strife, particularly when differences of expectation and cultural understanding come into play.

The **Formal** category was aimed at participating in larger society, mostly defined as large companies and government institutions. Low-literates were more negative about this level of interaction. Significant animosity was leveled towards these institutions, which were seen as either ignorant on low-literate issues ("*...they all always assume you can just understand what they write*") or actively malicious ("*If you're behind on your rent, they'll knock on your door, but if you need the housing association for repairs [...] then things start getting difficult*"). Low-literates experience a sense of stigmatization, which is made worse by an increasing demand that technology is used in participation activities. One NT1 participant explained that "*When I lost my job, I had to register myself for benefits [...] Well they give you this envelope, good luck, here's the computer. And you're just flabbergasted, because you can't do anything [...] they hand you this thing and hey, figure it out yourself.*" NT2 participants in particular did commend the rules-driven nature of Dutch society, which was considered easier to get into ("*I'm good at following Dutch rules, like separating garbage, going to school on time, staying polite*") and beneficial for society as a whole ("*In my country, there's discrimination. And no rules. But in the Netherlands, there are rules for... streets, and cars, and fishing. Very good, that's safe*"). The codes were further clustered into four second-order categories:

- *Norms And Values*: Overview of how societal expectations influence the participation experience of low-literates. For NT1 low-literates, Dutch expectations of self-reliance and self-efficacy are expressed in the common assumption that 'everyone can read and write', and that everyone can use computers and the Internet. NT2 low-literates encounter cultural differences regarding norms and values.
- *Barriers*: Those things done by societies and formal institutions that harm and hinder low-literates. Some barriers are attributes to the malicious behaviour of formal institutions, while other barriers relate to impersonal rules and laws that do not take low-literacy into account. This results in significant impediments to participation and learning, such as technology impediments, financial barriers, and a sense that no individual person inside formal institutions 'really cares' about low-literate issues.
- *Stigmatization*: Overview of how societal behaviors and expectations make low-literates feel stigmatized and 'looked down upon'. This contributes to a common sensation that low-literacy is 'invisible' in the Netherlands, and leads low-literates to develop ways of hiding their literacy issues.
- *Technology*: Overview of how the growing role of technology in modern-day societal participation impacts low-literates. The widespread and increasingly mandatory nature of technology and the rapid pace of technological chance present significant participation hurdles, usually overshadowing the benefits that technology can have for low-literates.

In the last round of clustering, researchers found that the focus of the remaining codes was on interaction between people and society. In these data, low-literates often spoke about their difficulties with certain participation behaviours, without associating these difficulties either with their own skills or with a related institution. For instance, one participant told how "*...you get this piece, and you run into this really difficult word... you completely lose track [...] you can keep that up for five minutes, and then...*" Another succinctly said that "*You can't do everything over the phone.*" A third participant suggested that "*I really think they should work with icons, because... with images, you can do so much more, instead of just having* *long swaths of text there."* We called this category **Interaction**. The codes were further clustered into two second-order categories:

- *Information-Communication Attributes*: Attributes that describe the specific way two parties interact. A message has a certain medium, exists in a certain modality, and might have unique restraints (for instance, a phone call is ephemeral and time-constrained, while a letter is a physical item that can be handled).
- *Learning Attributes*: Attributes that describe the learning process. This includes the roles of peers and teachers in shaping the learning experience, and the importance of having a 'correct' learning difficulty. Of all second-order categories, this category is likely the most situational one: had we studied low-literates not in a classroom setting but (e.g.) at work, or at home, this may not have surfaced. Nevertheless, for low-literates that are engaged with learning, the importance of peer support and good lesson plans is intrinsically connected to their participation experiences.

3.4.2. Specification Refinement

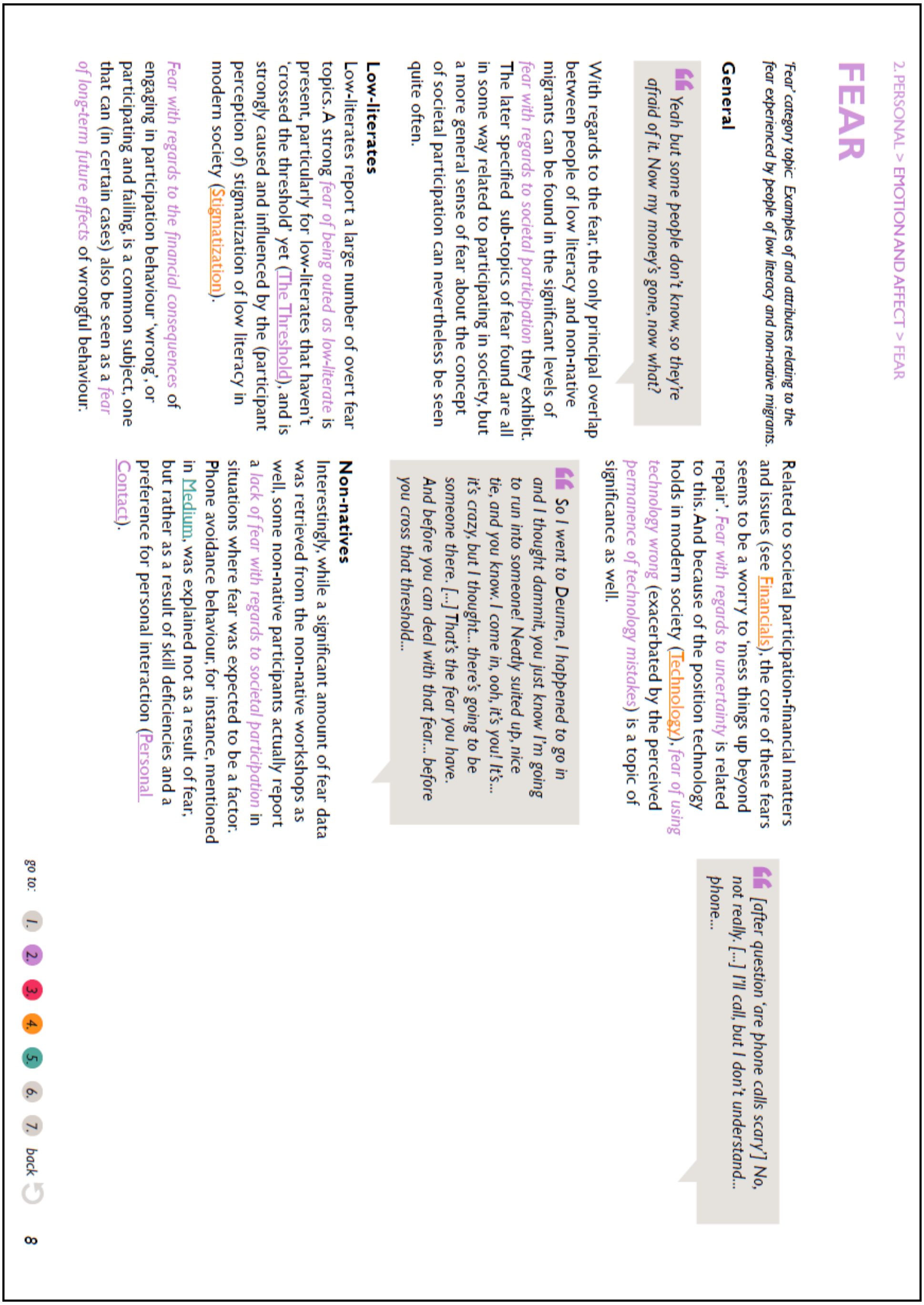
Using the SPELL model (Fig. 3.4), Schouten et al. [78]'s original VESSEL specification has been refined (cf. [73]). Empirical data was used to better situate the requirements in the daily life experiences of low-literates, using the four main categories as guidelines. This has resulted in longer, more detailed and concrete requirements. The requirements now combine a theoretical foundation of their concept with an empirical foundation of how this concept is useful specifically for low-literates. As a result, they are clearer, more actionable, and better matched to the low-literate societal participation problem area. For each requirement, the set of associated claims was analyzed. The intent was both to remove claims that are no longer seen as valid, and to add new claims that were initially overlooked. In practice, only new claims were added. These new claims describe empirically grounded expected effects of VESSEL. All existing claims were kept.

The refined specification is presented in Table 3.5. This table also shows which categories of the SPELL model (Fig. 3.4) have contributed to each of the new requirements and requirement-claim connections. A more expansive overview of the work involved in this, including direct links to the model's lowest-level nodes, can be found in the online supplement.

3.5. Conclusions

3.5.1. Research Question 1

This study's first question was "*How can the research methods 'participant workshop', 'cultural probe', and 'Grounded Theory' be used to map out the subjective societal participation experiences of low-literate citizens, and to update the operational demands with qualitative empirical data?*". This question is answered in sections 3.2-3.4. The combination of participant workshops and cultural probes was effective at collecting data with low-literate participants. The participant workshops



2.PERSONAL > EMOTION AND AFFECT > FEAR

FEAR

'Fear' category topic. Examples of and attributes relating to the fear experienced by people of low literacy and non-native migrants.

General

> *Yeah but some people don't know, so they're afraid of it. Now my money's gone, now what?*

With regards to the fear, the only principal overlap between people of low literacy and non-native migrants can be found in the significant levels of *fear with regards to societal participation* they exhibit. The later specified sub-topics of fear found are all in some way related to participating in society, but a more general sense of fear about the concept of societal participation can nevertheless be seen quite often.

Low-literates

Low-literates report a large number of overt fear topics. A strong *fear of being outed as low-literate* is present, particularly for low-literates that haven't 'crossed the threshold' yet (The Threshold), and is strongly caused and influenced by the (participant perception of) stigmatization of low literacy in modern society (Stigmatization).

Fear with regards to the financial consequences of engaging in participation behaviour 'wrong', or participating and failing is a common subject, one that can (in certain cases) also be seen as a *fear of long-term future effects* of wrongful behaviour.

Related to societal participation-financial matters and issues (see Financials), the core of these fears seems to be a worry to 'mess things up beyond repair'. *Fear with regards to uncertainty* is related to this. And because of the position technology holds in modern society (Technology), *fear of using technology wrong* (exacerbated by the perceived *permanence of technology mistakes*) is a topic of significance as well.

> *So I went to Deurne, I happened to go in and I thought dammit, you just know I'm going to run into someone! Neatly suited up, nice tie, and you know. I come in, ooh, it's you! It's... it's crazy, but I thought... there's going to be someone there. [...] That's the fear you have. And before you can deal with that fear... before you cross that threshold...*

Non-natives

Interestingly, while a significant amount of fear data was retrieved from the non-native workshops as well, some non-native participants actually report a *lack of fear with regards to societal participation* in situations where fear was expected to be a factor. Phone avoidance behaviour, for instance, mentioned in Medium, was explained not as a result of fear, but rather as a result of skill deficiencies and a preference for personal interaction (Personal Contact).

> *[after question 'are phone calls scary'] No, not really. [...] I'll call, but I don't understand... phone...*

go to: 1. 2. 3. 4. 5. 6. 7. *back* 8

Figure 3.5: Page from the interactive PDF, describing Personal Factors->Emotion and Affect->Fear. Three headers are shown. 'General' header describes the node as it applies to the whole low-literate demographic. 'Low-literates' describes node factors that apply only to NT1 learners, and 'Non-natives' describes factors for NT2 learners. Grey blocks with purple quotation marks give participant quotes (translated from Dutch).

		Inter-action		Formal				Informal			Personal						
Requirements	**Objectives**	Learning Attributes	Information & Communication Attributes	Technology	Stigmatization	Norms & Values	Barriers	Support	Social Consequences	Social Conflict	Social-Cultural	Self-efficacy	Needs	Motivation	Knowledge And Experience	Emotion And Affect	Cognitive Skills
R1. Adaptability: VESSEL should offer and/or support different learning styles and preferences. *The focus of adaptability should be on providing the right perceived level of difficulty. Exercises should be difficult enough to be useful, but not so difficult that they scare low-literate learners off.*	*O1.* O4. O5. *O8.*	X										X	X	X	X	X	X
R2. Sensitivity: 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.*	*O2. O3.* O5. O9.				X		X		X		X			X		X	
R3. Situatedness: 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. Learning exercises must teach low-literate learners to deal with cognitively, affectively, and socially challenging situations.*	O2. *O5.* O7. *O8.* O9.					X	X		X	X		X	X		X	X	
R4. Collaboration: 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.*	O2. O3. *O4. O5. O6.* O7. O9.	X		X	X	X					X	X	X		X	X	
R5. Multimodality: 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.*	O1. O2. O4. O5.		X			X							X	X		X	X
R6. Support: 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*	*O1. O2. O3.* O4. *O5.* O6. O7. O8. O9.	X					X	X					X	X		X	X
R7. Interactivity: 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.*	O4. O6. *O7. O8.*	X	X						X			X			X	X	
R8. Gaming principles: 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.*	*O3.* O5. O6. O9.										X			X		X	

Table 3.5: Refined VESSEL specification. First column provides expanded requirement descriptions: unformatted text is the original description, text in italics has been added. Second column shows claims associated with each requirement: unformatted claims are original, italic claims are new. Columns 3 to 18 show which elements of the SPELL model contribute to the requirement and claim updates, by way of an X in this column.

were valuable for collecting a variety of qualitative data. Recorded workshop audio was particularly useful: low-literate participants were eager to talk to peers and researchers, resulting in a large amount of rich data. By design, the data from these workshops mostly covered pre-selected topics of interest. The cultural probes were valuable for showing unexpected viewpoints. While the example behaviors used in the workshop were drawn from crucial practical situations (section 3.1, cf. [2, 3]) and earlier work with low-literates [106, 112], participants used the cultural probe to report many new behaviors and situations. Examples include 'driving a car and having to read road signs' and 'interacting with the elderly'. As outlined in the design, the workshops and probes taken together provided richness of data, answers to pre-existing questions, and new unexpected insights. This makes these methods suitable for mapping out the societal participation experiences of low-literate citizens in a comprehensive way. The data acquisition work done in this study has also been used by Cremers et al. [71] to provide a more general overview of lessons learned about using anthropological methods with special-needs demographics. The methods presented here have also been included in the Inclusive Design Toolbox [176], a collection of inclusive design methods aimed at supporting SMEs, as examples of how to approach a low-literate target audience.

Using Grounded Theory, we turned our rich empirical user experience data into the SPELL model (Fig. 3.4). The codes and categories used in the model were selected on quality of fit, not quantity of occurrence. This has allowed the broadest possible representation of the societal participation experience of low-literate citizens to make its way into the design process, letting us accurately adapt the VESSEL design to these experiences. As described in section 3.2.2, our shorter Grounded Theory method has some theoretical pitfalls that classic Grounded Theory does not have. In practice, we took steps to minimize the impact of this. Multiple researchers concurrently worked on the workshops, the transcription, and the final analysis, lowering the likelihood of significant researcher bias effects. As described, Kurvers et al. [89]'s language learner profiles were used to select participants, to ensure that all aspects of the low-literate societal participation experience were taken into account.

3.5.2. Research Question 2

This study's second question was "*Which VESSEL specification refinements are found by applying these methods?*" The refined specification in section 3.4.2 answers this question. The eight requirements have been expanded, and many new requirement-claim connections have been added (see Table 3.5). Empirical knowledge has been integrated into the specification, situating it more clearly in the real societal participation problems that low-literates experience.

The requirements have been expanded in three distinct ways. First, clear implementation and application directions were added to *all* requirements. Each requirement now better explains how it should practically be implemented in the context of VESSEL. Second, conceptual clarification was added to requirements R1, R2, R3, and R6. These requirements now clarify what idea they represent, and what the most important parts of that idea are for low-literate learners. Third, technical disclaimers were added to the requirements R4 and R8: for low-literate learners, a software implementation of 'collaboration' and 'gaming principles' is predicted to have potential drawbacks. It is unclear if the claimed benefits of these requirements outweigh those drawbacks.

17 new requirement-claim connections were added to the specification, bringing the total number up to 41. No claims were removed; all previously existing claims were confirmed viable by empirical data. All requirements now have 4 or more associated claims. Outliers are R4 (Collaboration, 7 claims) and R6 (Support, 9 claims). Requirement R6 in particular covers the entire Table 3.1 spectrum of learning effectiveness. This confirms the importance of learning support seen in literature, both in general [55] and specifically for low-literate learners [66]. Implementing support correctly – providing the right level of support, and giving learners the sense of being supported – should be a design priority.

3.5.3. SPELL Model

The SPELL model (Fig. 3.4) combines the theoretical insights and expert input of the societal participation axis model (Fig. 3.1, cf. [78]) with qualitative empirical data. The model advances our understanding of the experience of participating in society as a low-literate citizen by providing new data (based on lived experiences), and by structuring these data in a more insightful way. The added value of the SPELL model can be shown in three highlights. First, the 'social context' division of formal and informal social settings is supported by data: our results indicate that low-literate learners really experience the two types of social context as meaningfully different, as shown by the different cognitive, affective, and social problems they encounter. This corroborates literature expectations, where formal social settings (such as banking) are associated with comprehension problems and fear [82, 112], while informal social settings (such as participating in neighborhood activities) are associated with limited social inclusion and shame [1, 10]. The list of crucial practical situations explicitly lists formal and informal settings that pose a spectrum of possible cognitive, affective, and social problems (cf. [2, 3]); the SPELL model seems to confirm the value of this division, and highlights important details.

Second, the 'skill' division of information and communication skills is changed. Our data suggest that this division is too broad, and therefore not very useful for modeling or analysis. Instead, the 'Interaction' category in the SPELL model provides a detailed breakdown of the potentially difficult attributes of societal participation behavior in general. The higher level of detail makes this model more useful for analyzing challenging participation situations, and for describing low-literate issues in more applied, individualized contexts. We have used the NT1/NT2 division to provide more detail to the model, as this division is widely used in the Netherlands (cf. [4, 32]). It would also be possible to use Kurvers et al. [89]'s learner profiles, or any other meaningful division (ex. age, sex, schooling). The SPELL model can be used to describe these categories in more detail, by highlighting the attributes of interaction behavior that cause common issues or by suggesting solution directions. Description can even be done on the level of the individual learner.

The final category, 'Personal', also reflects the importance of a personalized approach, and provides detailed important areas for individualization. While the general added value of individualized learning and support is established (cf. [37, 60]), to the best of our knowledge, the SPELL model is the first model to provide a detailed overview of the lived experiences of low-literate citizens, and to contextualize this in the domain of societal participation learning. As a result, the SPELL model is a useful resource for design and development work with low-literate citizens. We ourselves will use it in the later development of VESSEL: the model will be used to write scenario-based use cases, to select learning content and practical exercises that match learner needs and learning objectives, and to ensure that the VESSEL system is evaluated in the proper context of use.

3.5.4. Limitations / Future Work

Practical issues in the application of the acquisition and analysis methods limit the study results. The participant workshop and cultural probe methods are difficult to implement optimally, and they are time- and labor-intensive: Participant workshops must be carefully prepared and supervised *in situ*, and cultural probes must be designed and assembled in collaboration with experts, to ensure a good fit to the participant demographic. Similarly, processing the results of the methods for later analysis – copying texts, transcribing audio, and cataloguing graphical results – is long and precise work. Given this, the ratio of 'useful' data to total data can seem low. While the unstructured, unsupervised nature of the methods is seen as a plus, some percentage of the results may not be valuable for the intended study goal (for example, one non-native participant reported on the trouble they had with adjusting their palate to Dutch cuisine). This should be factored into data acquisition planning. Finally, one unique cultural probe issue is that not all participants may return the materials on time, or at all. Particularly if quantity of results is a goal, other (combinations of) methods could yield results more quickly, or more consistently. The strength of these methods lies in providing rich data and unexpected insights, at the cost of high time and effort investment; whether or not they are useful for any given study should be carefully considered.

Our Grounded Theory results were limited by a relatively low total participant pool, which raises the question of whether or not the whole spectrum of experience was caught. Particularly in the technology use workshops, data came from only 16 participants. Furthermore, two obvious 'participant type' clusters were seen, matching the expected NT1/NT2 division: All native Dutch participants were relatively old, and from similar social backgrounds, while all non-native migrant participants were young adults or middle-aged, and had a variety of cultural origins and backgrounds. Our analysis treats all data equivalently, but this may not be valid. The SPELL model separates data by marking categories and nodes as being meaningful for NT1, NT2, or both (as seen in Fig. 3.5). A classical Grounded Theory analysis could take this division as a jumping-off point for later data gathering, but this did not happen in our study. Correspondingly, while the SPELL model remains the strongest way of describing the overall societal participation experience of low-literates, specialized models for the sub-demographics could yield even more useful results. Future studies should investigate this.

Limitations aside, this study presents clear venues for our future work. Now that the VESSEL system specification has been refined to be theoretically and empirically comprehensive, we can move to the prototyping and evaluation stage of SCE (Fig. 3.2). The high-level claims underlying the expanded requirements can be made specific and measurable by creating practical use cases (generated using the SPELL model). We can then translate the VESSEL design specification into functional prototypes and evaluate these with low-literate learners, to experimentally evaluate whether or not the claimed effects bear out.

4

First VESSEL Prototype

People of low literacy could benefit from automated support when learning about societal participation. We use the Socio-Cognitive Engineering method to design an Embodied Conversational Agent (ECA) 'coach' that can provide effective learning support to low-literate learners, to develop a prototype virtual learning environment consisting of an ECA coach and societal participation exercises, and to evaluate this prototype with low-literate end users. First, an inventory is made of the envisioned benefits of ECA coaching in the area of cognitive, affective, and social learning support. Second, existing high-level requirements are updated to better specify functional demands for the coach ECA, and for a corpus of societal participation learning exercises. Third, use cases are written, and used to develop the prototype. And fourth, the prototype is evaluated with low-literate users. In a mixed-method within-subjects experiment, low-literate learners complete sets of exercises with and without support from the coach. 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. These caveats give direction for improvement and future study.

The content in this chapter has been submitted for publication as: Schouten, D. G. M., Deneka, A. A., Theune, M., Neerincx, M. A., & Cremers, A. H. M. (under review). An Embodied Conversational Agent Coach to Support Societal Participation Learning by Low-Literate Users.

4.1. Introduction

In modern information societies (such as the Netherlands) the societal participation of low-literate citizens is problematically low [4]. 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. [78, 109]) or social (low motivation and desire to learn, or low trust in teachers and other learners, cf. [1, 10]). 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* [2, 89]. To this end, we are designing the virtual learning environment VESSEL: A Virtual Environment to Support the Societal participation Education of Low-literates [78]. 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. [34–36]), *learning experience* (the learners' skills, needs, and wishes should be incorporated throughout the learning, cf. [36, 159, 162]), and *learning outcomes* (the learning should aim to reach meaningful and desired goals, cf. [34, 158–160]) [156]. 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: 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. 4.1) 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 [84, 85, 163]. 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. [177]). This specification can then be developed into a prototype, which is used to experimentally evaluate the claims.

Earlier work [78, 156] has resulted in a high-level VESSEL specification, consist-

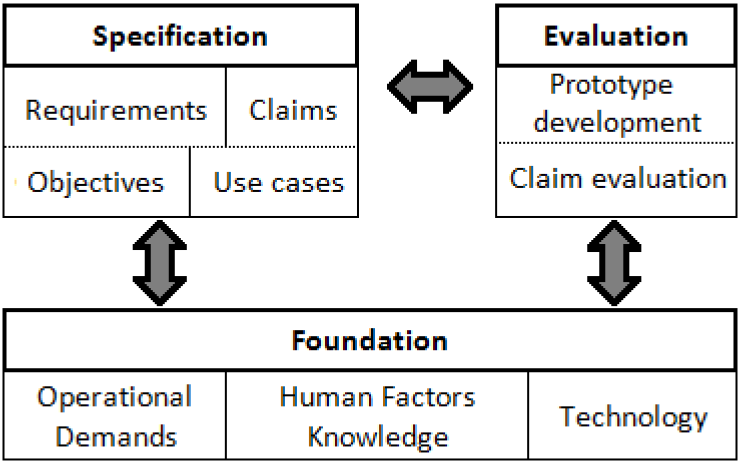


Figure 4.1: Socio-Cognitive Engineering method [84, 85].

ing of a requirements baseline with eight requirements (see Table 4.1), and claims that connect these requirements to the nine system objectives of learning effectiveness. This requirements baseline is theoretically supported [78]: The requirements in the specification were derived from theories of adult learning (andragogy [123–125], transformative learning [126, 178], constructivism [23, 24], and e-learning [134–136]) and theories on computer-supported learning that highlight the value of information provision, worldwide communication, interactivity, and use of gaming principles / gamification [56]. The baseline is also empirically grounded [156]: The requirements were refined by applying Grounded Theory [74] 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. [72, 164]) 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. 4.1). 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. Fig. 4.2 presents the envisioned

VESSEL system setup, showing the ECA coach and exercise elements. Here, exercises are scenario-based training (cf. [97]) situated in crucial practical situations important to low-literate learners. ECAs are "*anthropomorphic interface agents*" [40, 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: ECAs can adapt their looks and behaviours easily, allowing 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 [156], 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 evaluate 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. [179]) 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 [156, 180]. Additionally, Kramer et al. [181] 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. [182] 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 is a unique and interesting contribution. A complementary study by Schouten, Venneker et al. [183] 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 cog-

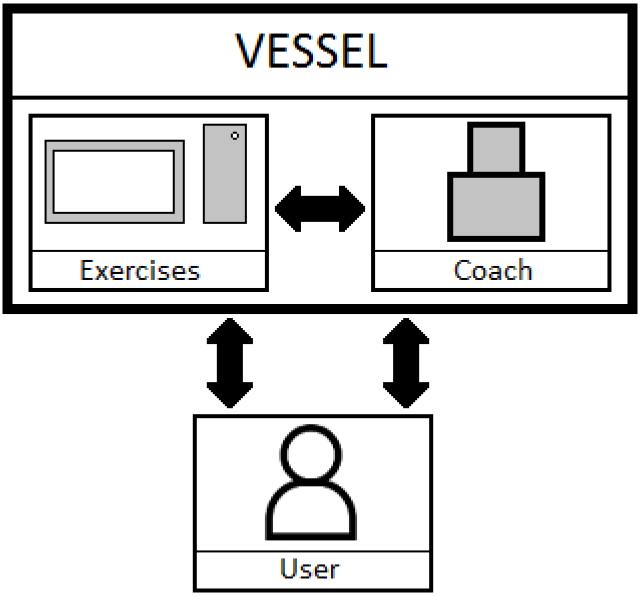


Figure 4.2: VESSEL system setup. Three double-sided arrows indicate user-system interactions. Bottom left arrow: The user performs exercises. Bottom right arrow: The coach monitors the user's actions, and interacts with the user by giving feedback and support. Top arrow: The coach monitors exercise state and changes support as appropriate.

nitive, 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. 4.1). 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. And 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 4.2 shows the updated VESSEL foundation. In section 4.3, the specification requirements are refined, and use cases are derived and written. Sections 4.4, 4.5, and 4.6 show the evaluation process. In section 4.4, the prototype is described in terms of functionality, interaction methods, and appearance. Section 4.5 describes the experiment created to evaluate the effectiveness of the learning support provided by the prototype's ECA coach. Section 4.6 presents the results of the evaluation. Finally, section 4.7 presents conclusions, discussion of findings, and directions for future work.

4

4.2. Foundation

4.2.1. Technology: Embodied Conversational Agents

Embodied Conversational Agents (ECAs) are a subclass of Intelligent Virtual Agents: Autonomous software programs that can interact with humans and other agents [40]. 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 [42, 184, 185], attractiveness [43, 186], and cultural appearance stereotypes [184, 187, 188]. In addition, ECAs can use not only verbal communication behaviours (e.g. speech and natural language understanding), but also non-verbal behaviours (e.g. body language, gesturing, facial expressions, and gaze direction [41, 181, 182]). 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 [40, 41, 189, 190]. This lets humans react to the social cues and behaviours of ECAs as if they were human conversation partners [44–47].

Because ECA behaviours and appearances can be adjusted [22, 57], ECAs can be used to fulfill a variety of roles in a virtual environment. For instance, Bickmore et al. [49] 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. [48] 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. [191] and de Rosis et al. [192]

were effective in changing user behaviour patterns. Shamekhi et al. [193] show that an ECA coach focused 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 [194] 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. [181] 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 [191] health counselor increased physical activity and fruit and vegetable consumption, and Veletsianos and Miller [195] 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 [196]. De Rosis et al. [192] show that users converse with and learn from an ECA for health promotion with adaptive dialogue, and Miao et al. [197] 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. [198] 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*" [199, p. 6]; they call this the *persona effect*. And socially, digital coaches can be seen as 'friends' and trusted mentors in a learning system [192, 200], forming a long-term relationship of trust between learner and coach [201, 202]. Ter Stal et al. [182] 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. [203]). ECA digital coaches can use nonverbal behaviours and appearance factors, such as similarity attraction, to form these relationships quickly and strongly [42]. Digital coaches also enhance engagement and learning in a virtual learning space, by acting as conversation partners that human users will genuinely talk to [195].

4.2.2. 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 [204]. This way, "*students are encouraged to develop their own creativity, motivation, and resourcefulness*" [93, p. 652]. The coach can use verbal scaffolding techniques [205, 206] 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 counseling technique aimed at leveraging intrinsic motivation to enact behavioural change [207]. The motivational interviewing techniques help learners to feel good about the learning process, and to reframe and solidify positive self-efficacy information (cf. [208–210]). The coach can use motivational interviewing techniques by offering exercise-specific feedback, eliciting self-reflection, and applying social persuasion to raise learner self-efficacy [211]. Social support is operationalized as *small talk*, which is a cornerstone of building trust. Trust is an important element of the learning process [40, 212], 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' [153]. The coach can apply these categories in exercise-specific small talk.

4.2.3. 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: One, 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. And two, 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 [156]. 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. [156] 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' [213] overview of critical online banking usability factors, and Nielsen's [214] and Leavitt & Schneiderman's [215] 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 infor-mation-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. 4.3.

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 [216] overview of social demeanor 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 [217], 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; literature currently does not show clear consensus on which ECA designs are preferred in which situations [182, 218], 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 [42]. ECAs were drawn from Brinkman et al.'s [58] 'Virtual Reality Exposure Therapy (VRET)' virtual environment. Participants were given paper pictures of a service desk setting (Fig. 4.4) and 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 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,

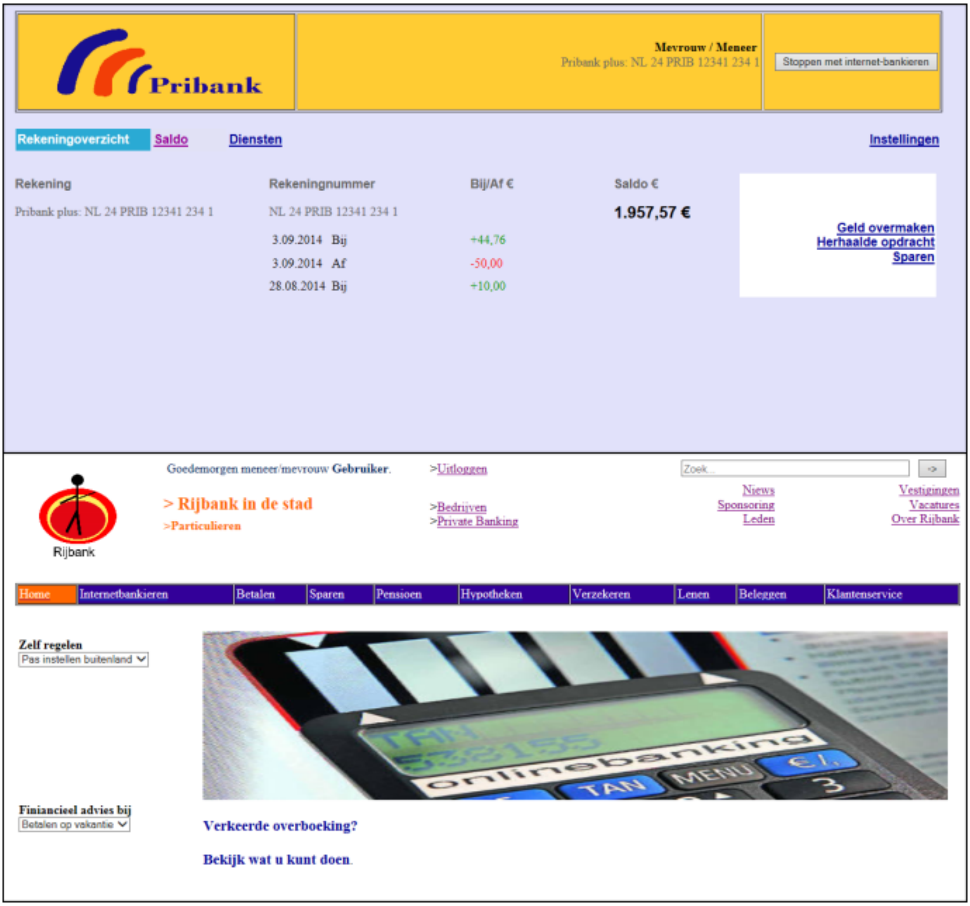


Figure 4.3: Online banking exercise websites. Left: Easy Online Banking. Right: Hard Online Banking.

swaying left and right slightly while sitting on a chair; apart from this, they employ no other body movement or gestures of any kind. Both ECAs can be seen in the context of the Service Desk exercises in Fig. 4.4; 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 minutes) and 'fail' (the exercise is completed incorrectly or not within 6 minutes) 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 time to reasonably

Figure 4.4: '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.

do the exercises, while a longer limit would inflate the time footprint of the envisioned experimental study (see Section 4.5).

4.3. Specification

Following Fig. 4.2, 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 [73]. 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 4.1, below, shows the new requirements baseline. An in-depth description of the refinement process can be found in Appendix B. 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 C.

4.4. Evaluation: Prototype Development

Functionality. The prototype consists of the ECA coach that offers cognitive, affective, and social learning support as described in sections 4.2.1 and 4.2.2, and the exercises described in section 4.2.3 (*Easy* and *Hard Online Banking*, and *Easy* and *Hard Service Desk*). Cognitive support is offered *during* the exercises. Bloom's [219] 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

General Requirements	Coach Requirements	Exercise Requirements
R1. Adaptability. 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.	**R1.1-C.** 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'.	**R1.1-E.** 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.
R2. Sensitivity. 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.	**R2.1-C.** The coach should always address learners calmly and kindly, and avoid using phrases and broaching topics that upset low-literate learners.	**R2.1-E.** The exercises should be as sensitive as needed to reach the intended learning goals and difficulty level.
R3. Situatedness. 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.		**R3.1-E.** 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.
R4. Collaboration. 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.		
R5. Multimodality. 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.	**R5.1-C.** The coach should combine audio 'speech' with visual and textual supporting material.	**R5.1-E.** The exercises should be as multimodal as needed to reach the intended learning goals and difficulty level.
R6. Support. 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.	**R6.1-C.** The coach should use verbal scaffolding techniques to offer cognitive learning support. **R6.2-C** The coach should use motivational interviewing techniques to offer affective learning support. **R6.3-C** The coach should use small talk to offer social learning support.	
R7. Interactivity. 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.	**R7.1-C.** The coach should interact with users actively by starting conversations and offering help. **R7-2-C.** The coach should react to learner questions and demands for help.	**R7.1-E.** The exercises should be interactive, requiring learners to use input mechanics to engage with the virtual environment in order to complete them.
R8. Gaming principles. 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.	

Table 4.1: Refined VESSEL requirements baseline, from the perspective of VESSEL as an ECA coach supporting exercises. Column 'General Requirements' shows the high-level requirements that were created in [156]. Columns 'Coach Requirements' and 'Exercise Requirements' show zoomed-in requirements created in Appendix B

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 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 [179]. Both ECA programs 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.

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 minutes vs. the exercise was completed in 3 to 6 minutes 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.

Figure 4.5: 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.'

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 have been shown in section 4.2.3 (see Figures 4.3 and 4.4). The coach ECA's visual appearance was based on the same pre-study used for the conversation partner ECAs [217]. 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. 4.5). Like the conversation partner ECAs, the coach ECA has one facial 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.

4.5. Evaluation: Methods

4.5.1. 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 C). 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.

4.5.2. Measures

Twenty dependent variables were measured: Eighteen variables were self-report measures, obtained using two questionnaires (see section 4.5.4), and two variables were objective performance metrics. Table 4.2 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?

4.5.3. Participants

Twelve low-literate people participated in the entire experiment. Kurvers et al.'s [89] 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,

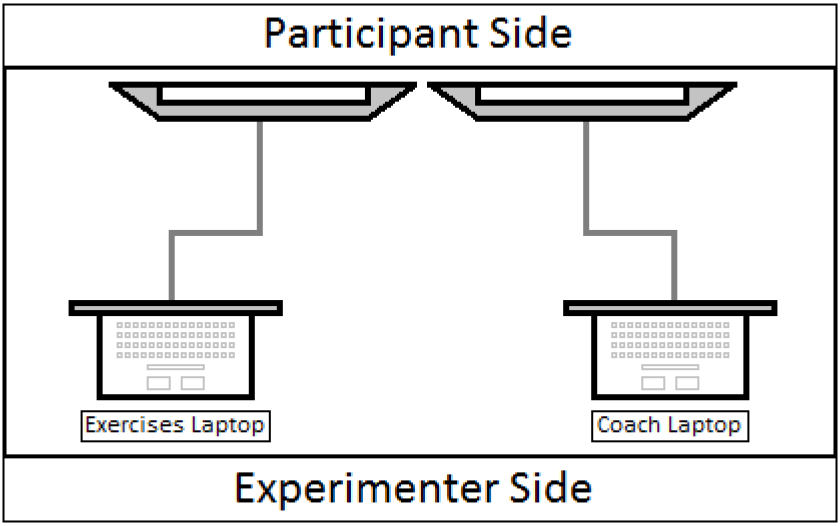


Figure 4.6: Schematic overview of experimental setup: 2 laptops (lower figures) connected to 2 monitors (upper figures). Monitors are placed and angled such that participants could not see the laptops and the experimenters while seated.

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 [217].

4.5.4. Materials

The experimental setup consisted of two laptops connected to two additional monitors (Fig. 4.6). The laptops, on one side of the table, were used by the experimenters. The monitors, on the other side 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 4.2): 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. 4.7). Participants were told to mark one of the five boxes per question. Bars were labeled '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

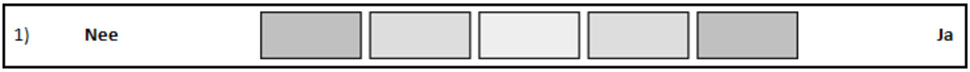


Figure 4.7: Example answer bar for the short and SQs ('Nee' means 'No', 'Ja' means 'Yes').

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.

4.5.5. Procedure

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

Variable	Hypothesis	Description
Subjective Measures: Exercise Questionnaire (EQ)		
EQ.1 Perceived performance (exercise)	H1	"I have done the exercise about (*online banking / talking to people at a service desk*) well."
EQ.2 Difficulty (exercise)	H5	"I found the exercise to be difficult."
EQ.3 Self-efficacy (exercise)	H5	"I am now better at (*online banking / talking to people at a service desk*)."
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."
Subjective Measures: Session Questionnaire (SQ)		
SQ.1 Positive affect (language class)	(*practice question*)	"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."
Objective Measures: Measured Per Exercise		
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.

Table 4.2: Overview of quantitative measures. Includes measure source (EQ, SQ, or direct measurement), applicable hypothesis, and question wording or measure description.

4

4.6. Evaluation: Results

Three sets of analyses were done. 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. And 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 α =.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 α =.600.

4.6.1. 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'. And 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 4.3 shows means and standard deviations of the five questions for each of the eight measurement moments. Table 4.4 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 Banking, but higher experienced difficulty for Service Desk; Without Coach did not show this.

	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
	.75	*1.54*	*.52*	*1.54*	*1.68*	*1.17*	*.98*	*1.14*
EQ.2 Experienced difficulty (exercise)	-1.25	.92	-1.50	-.75	-.42	1.58	-1.67	-1.33
	.97	*1.00*	*.90*	*1.48*	*1.68*	*.90*	*.89*	*1.15*
EQ.3 Self-efficacy (exercise)	.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*
EQ.4 Positive affect (exercise)	1.50	.75	1.58	1.41	.92	-.50	1.67	1.42
	.80	*1.50*	*.51*	*.67*	*1.62*	*1.57*	*.65*	*.67*
EQ.5 Computer support	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*

Table 4.3: EQ means (*standard deviations*). Mean scores range from [-2,2].

4

- **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. And 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.

4.6.2. 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 4.5, below, shows the factor loadings for this solution.

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

Table 4.4: Significant results of EQ repeated-measures GLM analysis. For each question and each factor / group of factors, F-value (F), significance (p), and observed power (β) are given if p<0.05.

	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

Table 4.5: Factor loadings for the 11 questions used in the factor analysis. Only factor loadings of .500 and higher are shown.

Based on the factor loadings shown in Table 4.5, a two-factor solution was decided on. Factor 1, 'Information Skills', contained questions 2, 5, 6, and 10, with a reliability of α=.86. Factor 2, 'Communication Skills', contained questions 3, 4, 9, and 11, with a reliability of α=.82. While questions 7 and 12 seemed to form a third factor, the reliability of this factor was only α=.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 seen: 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 order. Table 4.6 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 4.6: 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. 4.8. Before the first experimental session and before and after the second session, 'Communication Skills' was significantly higher than 'Information Skills'.

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

Table 4.6: SQ data means (*standard deviations*). Mean scores range from [-2,2].

	With Coach		Without Coach	
	Easy	Hard	Easy	Hard
DM.1 Completion time (s)	280.2 *128.3*	364.8 *113.4*	209.3 *107.2*	347.9 *77.8*
DM.2 Completion rate	.83 *.39*	.50 *.52*	.75 *.45*	.17 *.39*

Table 4.7: Performance metrics means (*standard deviations*) for Online Banking exercises. Completion time ranges from [0-600] (in seconds). Completion rate ranges from [0-1].

4.6.3. 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 was 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 strong-ly 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 4.7 shows means and standard deviations.

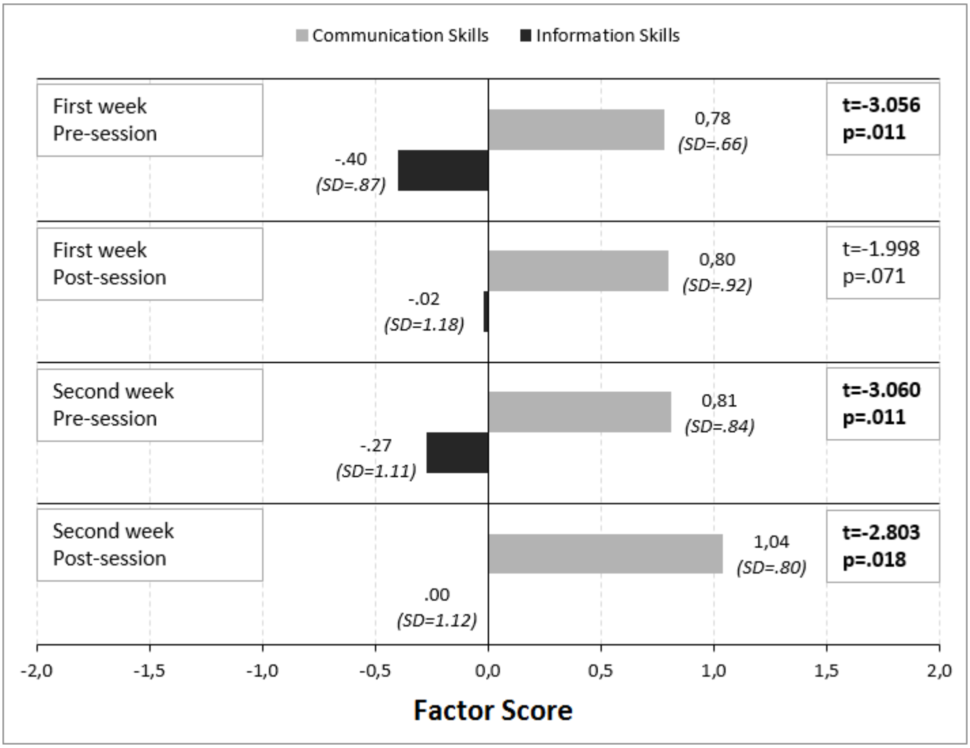


Figure 4.8: Means for 'Information Skills' and 'Communication Skills' factors, across the four measurement moments. Values next to bars represent mean (standard deviation). Boxes on the right show the test statistics of the paired-samples T-test that compared means for Information Skills and Communication skills in that measurement moment; bold text indicates significant results.

4.6.4. 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. But 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 programs 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. 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'.

4.7. Conclusions and Discussion

4.7.1. Findings

Building on earlier design work for the envisioned system VESSEL [78, 156], 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, 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 4.1). 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?*", was answered in section 4.2. 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?*" was answered in sections 4.3 and 4.4. 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 [217] 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 [42, 43]. But instead, all ECAs in the prototype were valued on matching the (visual) stereotype of their role. Participants chose the Fig. 4.5 coach because 'she looked friendly and approachable', and the Fig. 4.4 *Easy Service Desk* conversation partner 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 [42], but does confirm that stereotype-reinforcing appearances can have a strong impact [187]; our results seem to suggest a 'job-appropriate clothing' stereotype rather than Angeli & Brahnam's [187] 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 [186] and Nass et al. [43]. 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 section 4.6, 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 [48] ECA coach, and similar improvements in the affective experience are reported by Lester et al.'s *persona effect* study [199], Lane et al.'s computer science education ECA coach [48], Bercht & Vicari's pedagogical support agent [220], and Shaw et al.'s embodied situated support agent [198]. 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 [191], and increasing learner involvement in dialogue with a learning agent [192]) and user-system satisfaction [49]. 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.

4.7.2. 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). But 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. But 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-centered design and testing could ensure that the different ECAs are actually perceived as uniquely meaningful. And second, the 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. But 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 [179]. During the small talk and motivational interviewing sections, there was a flow structure based (partially) on measurable objectives and keywords. But 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). And interaction-wise, 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 non-verbal 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 [195] 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 [221] 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 [2, 80, 81, 89]. 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.

4.7.3. 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 section 4.1 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 [202], a friendly relationship of trust has started to form. Miao et al. [197] 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 [89] profiles 2 and 3, see section 4.5.3), 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.

5

Formalized Cognitive Learning Support

In this study, we attempt to specify the cognitive support behaviour of a previously-designed Embodied Conversational Agent coach that provides learning support to low-literates. Three knowledge gaps are identified in the existing work: An incomplete specification of the behaviours that make up 'support', an incomplete specification of how this support can be personalized, and unclear speech recognition rules. We use the Socio-Cognitive Engineering method to update our foundation of knowledge with new online banking exercises, low-level scaffolding and user modeling theory, and speech recognition. We then refine the design of our coach agent by creating comprehensive cognitive support rules that adapt support based on learner needs (the 'Generalized' approach), and attune the coach's support delay to user performance in previous exercises (the 'Individualized' approach). A prototype is evaluated in a three-week within- and between-subjects experiment. Results show that the specified cognitive support is effective: Learners complete all exercises, interact meaningfully with the coach, and improve their online banking self-efficacy. Counter to hypotheses, the Individualized approach does not improve on the Generalized approach. Whether this indicates suboptimal operationalization or a deeper problem with the Individualized approach remains as future work.

The content in this chapter has been published as: Schouten, D. G. M., Massink, P., Donker, S. F., Neerincx, M. A., Cremers, A. H. M. (2020). Using scaffolding to formalize digital coach support for low-literate learners. User Modeling and User-Adapted Interaction (2020) [222].

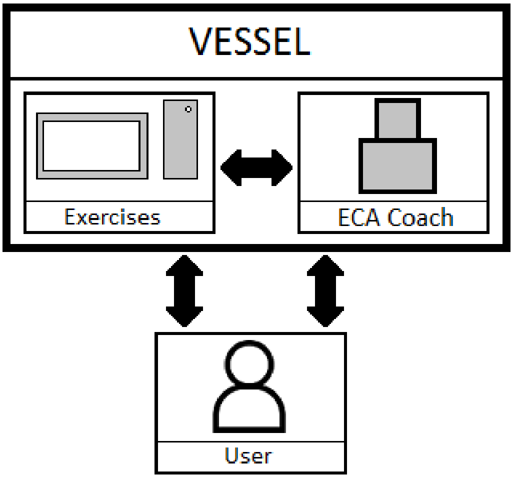


Figure 5.1: Envisioned VESSEL design. Arrows indicate system interactions: The user performs exercises, the ECA coach monitors exercise state and user-system interaction, and the coach supports the user as appropriate. Image from Schouten et al. [223]

5

5.1. Introduction

People of low literacy struggle to independently participate in information societies [4]. Limited information (reading and writing) and communication (speaking and understanding) skills lead to participation issues, which can be cognitive, affective, or social in nature [78]. Cognitive issues relate to applying information and communication skills, and possessing general knowledge about society. Affective issues relate to fear, shame, and low self-efficacy. Social issues relate to lack of motivation and trust in others. These issues can be addressed by providing societal participation learning that is grounded in *crucial practical situations* (real-life participation scenarios that involve the skills and knowledge needed to participate in society independently, such as online banking, grocery shopping, or engaging with local government; cf. [2, 3]), which allows low-literate learners to practice skills and gain knowledge and experience in a practical context of use. For this learning to be *effective*, especially for learners with limited information and communication skills, such as low-literate learners, the learning must be accessible (barriers to entry are lowered or removed), the learning experience must be positive (learners can and want to engage with the learning), and learners must reach desired learning outcomes [156]. We aim to provide effective learning with VESSEL: A *Virtual Environment to Support the Societal participation Education of Low-literates* [78, 156, 223]. VESSEL consists of situated, interactive exercises in the societal participation domain, and an autonomous, rules-driven Embodied Conversational Agent (ECA) coach that supports low-literate learners before, during, and after these exercises with cognitive, affective, and social learning support (see Fig. 5.1).

We use the Socio-Cognitive Engineering method [SCE, cf. 84, 85, 163] in the

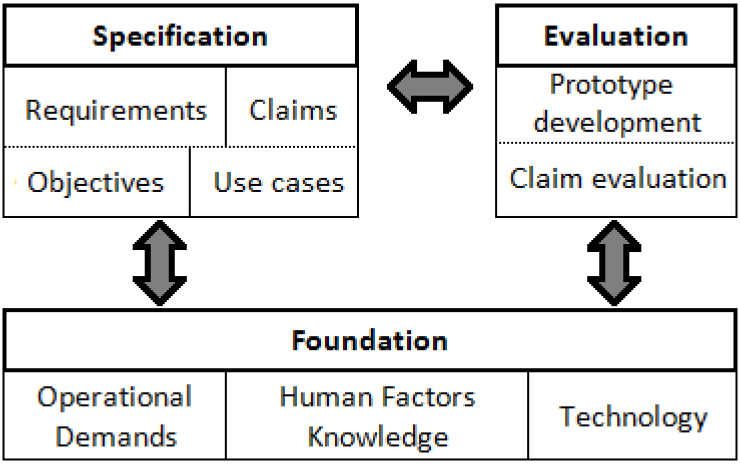


Figure 5.2: Socio-Cognitive Engineering method used in this study. Double-sided arrows between the Foundation, Specification, and Evaluation boxes indicate that development can move to any phase at any time [84, 85, 163].

development of VESSEL. The SCE method is an iterative software design and development method that moves (non-linearly) through three phases, shown in Fig. 5.2. In the *foundation* phase, relevant operational demands (the software system's context of use), human factors data (theory relevant to user-system interactions), and technology (both technology currently in the system and envisioned technology) are combined into a foundation of data. In the *specification* phase, a requirements baseline is created containing requirements, claims, system objectives, and use cases. This is then used for the *evaluation* phase, where the validity of the specification is empirically tested. Evaluation results are used to iteratively update the foundation and refine the specification.

Previous work used a high-level requirements baseline (see Table D.1 in Appendix D) to develop a first VESSEL prototype, consisting of an ECA coach that offered three kinds of learning support for four exercises [easy and hard 'online banking' and 'service desk conversation' exercises, cf. 223]. Cognitive support based on scaffolding, a teaching method that provides the right level of support at the right time [204], was offered during the exercises. Affective support based on motivational interviewing, a counseling technique that focuses on behavioural change [207], was given after the exercises. Social support based on small talk, a form of social interaction important for building trust [153], was used before the exercises. All support was provided in the form of pre-recorded spoken utterances and controlled by an operator, using the Wizard-of-Oz method to act as an ECA behind the scenes [cf. 179]. Notably, support was both created and provided in an *informal* manner. Support utterances were created based on an expert walkthrough of the system: Researchers determined areas where low-literates would likely struggle, and wrote utterances to address the predicted issues. And during the exercises, the Wizard-of-Oz operator interpreted user actions and speech and selected the utterance(s) considered best in this situation. Evaluation showed that the ECA coach resulted in a more positive cognitive, affective, and social learning experience, and higher self-efficacy about difficult online banking scenarios. As proof-of-concept, this shows that VESSEL can improve learning effectiveness for low-literate learners.

As the results from Schouten et al. [223] were promising, the next development step is to create a *formal* design specification that accurately describes VESSEL's envisioned functionality as automated learning support. This involves two things: First, writing a comprehensive set of dialogue rules for the ECA coach's cognitive, affective, and social support behaviour, that can be applied by automated computer support without requiring human interpretation. And second, incorporating new functionality as needed to improve support provision and learning effectiveness. Each of the three support types need a separate refinement step. We focus on the coach's *cognitive* support in the present study, as effective cognitive support is necessary to ensure learners can understand the system and complete exercises. Affective and social support are left to later work.

Our current implementation of cognitive support has three relevant knowledge gaps which the formalization process must address. First, because the existing set of coach support utterances is based on a non-comprehensive expert walkthrough, the utterances do not yet structurally and comprehensively cover the exercises. Not all challenging exercise elements have associated support utterances, and the existing utterances contain different levels of information and direct guidance, with no clear underlying logic. Formalized support will require a comprehensive set of support utterances for each exercise, in which the utterances cover every relevant aspect of the exercise and in which they are comparable in terms of information provided. Second, the coach's speech recognition functionality requires further operationalization. As the current speech recognition is left up to the Wizard operator's interpretation of user utterances and context, there are no formal rules in place to specify what learner utterances the coach should react to, and how. Formalized support will require a clear, unambiguous speech recognition ruleset. Third, we expect that *personalising* cognitive learning support will substantially improve learning outcomes. But our current implementation of cognitive support does not have a coherent and unequivocal specification of *how* this support can be personalized. We hypothesize that (in concert with the above) VESSEL's learning effectiveness could be improved by incorporating *user modeling* [the process by an intelligent system infers user traits from user-system interaction, cf. 37, 224–226] to better adapt the offered support to individual learners' circumstances and needs. To achieve this, formalized support will require a clear user model of support need, including an unambiguous list of user actions relevant to this model and a description of changes to the coach's support provision over time that can be made on the basis of this.

In this work, we aim to design and evaluate a VESSEL prototype that offers formalized cognitive learning support. Four steps are needed. First, we update the VESSEL foundation in three ways. We update operational demands by designing exercises based on crucial practical situations that demand cognitive support. We update human factors knowledge by incorporating more detailed scaffolding theory, as well as theory concerning user modeling. And we update technology by describing the envisioned role of speech recognition. Second, we refine the VESSEL specification: We operationalize the foundation theory into a comprehensive set of coach dialogue rules, update the requirements baseline, and write a use case to illustrate expected findings. Using the refined specification, we define in what ways the coach can provide cognitive support based on the learner's progress in the current exercise. We call this approach to support provision the 'Generalized' approach. We also describe how the coach models the learner's skill level based on their performance, and how it can use this model to attune its support provision in later exercises. We call this the 'Individualized' approach. Third, we design and develop a VESSEL prototype, consisting of an ECA coach that can offer cognitive learning support along both the Generalized and Individualized approaches, and three online banking exercises. This prototype will be designed for use in a Wizard-of-Oz experimental setup, in which an operator applies the coach's support behaviour and speech recognition behind the scenes by selecting prescribed outputs for the computer-sensed inputs [179]. Fourth, we experimentally evaluate the prototype with low-literate learners. We investigate how the new prototype affects the cognitive, affective, and social learning experience and learning outcomes, compared to our previous work, and we investigate if using both the Generalized and Individualized approaches leads to higher learning effectiveness than only using the Generalized approach. This leads to the following research questions:

- **Q1. Design.** How can we create a formal design specification for VESSEL that incorporates rules for cognitive learning support provided by an ECA coach?
  - **Q1a.** Which operational demands, human factors knowledge, and technologies are needed to write these rules?
  - **Q1b.** Which functionalities, interaction methods, and appearances should the ECA coach have to reflect this specification?
- **Q2. Evaluation.** What is the learning effectiveness impact of a VESSEL prototype that offers cognitive learning support according to the formal specification?
  - **Q2a.** Are the learning effectiveness results of this prototype comparable to the VESSEL prototype that offered informal cognitive, affective, and social learning support?
  - **Q2b.** Does using both the Generalized and Individualized approaches to learning support result in higher learning effectiveness than using only the Generalized approach?

The structure of this paper is as follows. Section 5.2 provides the refinement of the sCE foundation, necessary for deriving the concrete design specification in section 5.3. Section 5.4 describes the resulting new VESSEL prototype. Sections 5.5 and 5.6 describe, respectively, the experiment that evaluates the prototype, and the evaluation results. Section 5.7 presents conclusions and directions for future work.

5.2. Foundation

5.2.1. Operational Demands: Exercises

To accurately evaluate the effectiveness of cognitive learning support, exercises are needed that pose a significant cognitive challenge and demand coach support, but that can be completed with this support. If the exercise is too easy, learners will not require support; if the exercise is too difficult, no level of support will be effective. The first VESSEL prototype [223] contained four exercises: An easy and a hard exercise about online banking, and an easy and a hard exercise about visiting a government service desk. Of these, only the hard online banking exercise meets our needs: The exercise was challenging and demanded significant coach support, but participants often completed it. For this prototype, three new challenging online banking exercises were created, using the 'Hard Online Banking' website from Schouten et al. [223] as a task environment. In Exercise 1, the user must transfer money from their checking account to a webshop. In Exercise 2, the user must report a change of address to their bank. In Exercise 3, the user must transfer money from their savings account to their checking account. All exercises are intended to be equivalently challenging. To achieve this, we ensured that each exercise had the same number of *critical waypoints*, which we defined as those exercise steps that a learner *must* take to successfully complete it. In the context of online banking, critical waypoints can either be *navigation waypoints* (getting to the right part of the online banking website at the right time) or *data entry waypoints* (entering the right information in the right place). Each exercise was designed with exactly four navigation and four data entry waypoints, presented in the same order: Three navigation waypoints, then four data entry waypoints, then one last navigation waypoint. All exercises come with written summary instructions showing the goal and necessary information, such as bank account number and money amount to transfer, or street name and postal code of a new address.

5

5.2.2. Human Factors Knowledge: Scaffolding

Three core elements of scaffolding are contingency, fading, and transfer of responsibility [227]. *Contingency* refers to matching support to the learner's current ability. Three types of contingency are identified: Domain contingency, instructional contingency, and temporal contingency. *Domain contingency* means ensuring that the exercise or (sub)task has the right level of challenge for the learner. Exercise challenge level should fall in the Zone of Proximal Development [228, 229]. Mislevy et al. claim that: "*...the most accurate information about a test taker is obtained when the level of difficulty is close to the test taker's level of performance. However, there is also an important experiential aspect (...) Items that are too hard demoralize the test taker, while items that are too easy bore her.*" [60, p. 112]. In VESSEL, we use exercise design to aim for domain contingency, as shown in section 5.2.1.

Instructional contingency refers to tailoring the amount of support to the learner's skill level. This is derived from constructivist views of learning, which claim that learners actively construct knowledge and meaning by interacting with their environment [23, 24]. Learners should complete as much learning by themselves as possible for optimal outcomes [204, 230], and they should attribute success to themselves instead of external sources, as this raises self-efficacy [109]. Support should not take over too much responsibility too quickly. In VESSEL, we reach instructional contingency by categorizing the coach's support utterances into two categories: *Proactive* and *reactive* utterances. The coach can use *proactive utterances* when it detects that the learner needs support (e.g. by observing that learners have not made progress for some time). This is necessary because learners in tutoring sessions often do not actively ask for help [231, 232]. We use van de Pol, Volman & Beishuizen's [227] overview of scaffolding tools to define five proactive utterance subcategories: A proactive utterance can be a *prompt* (a simple question to gauge the learner's knowledge level), an *explanation* (an answer to either an earlier prompt or a learner question), a *hint* (an implicit suggestion of what the learner should do next that references the correct next step), an *instruction* (an explicit description of what the learner should do next), or *modeling* (an offer to demonstrate what the learner should do next, followed by the coach actually demonstrating it). Each of these utterance types provides support at a different level of directness. We define *support level* as a measure of the amount of direct guidance in a support category; support levels go from 1 (prompt) to 5 (modeling) as shown in Table 5.1. The coach can use *reactive utterances* to respond to learner speech or actions (described in detail in section 5.2.4). Finally, the coach can give feedback based on learner progress. If the learner attempts to move to the next exercise waypoint and has taken all necessary steps *correctly*, the coach uses praising feedback; if the learner has taken any steps *incorrectly*, the coach uses corrective feedback to indicate that something went wrong. See Table 5.1.

Temporal contingency describes that support should be given at the right time, when the learner is confused or questioning [229, 233]. If support is provided too late, learners are frustrated by a lack of progress; if it comes too quickly, learning is impaired [230] and learners might resent the support for giving an answer they could have found themselves [234]. In VESSEL, we reach temporal contingency by defining when the coach should use support utterances. For proactive utterances, we define that the coach should wait a certain amount of time between utterances (to avoid information overload and give learners a chance to parse and react to the utterance): We call this amount of time the *support delay*. We set a support delay of 20 seconds based on timing analysis of our previous work [223]. Reactive utterances should be used as soon as the appropriate conditions are met, in order to be useful [235].

Fading refers to gradually lowering the amount of offered support over time, as the learner's skill improves. Traditionally, human tutors use scaffolding by setting difficult exercises and immediately providing 'heavy' scaffolding [quick proactive guidance with a high support level, cf. 236], and then lowering that heavy scaffolding as learners start performing better. However, previous work has shown that low-literate learners have strong negative emotional reactions to unexpected challenge and to exercises that exceed their self-confidence and self-efficacy [156]. A system that starts out with heavy challenge and heavy scaffolding may lead to learners 'giving up', and either quitting the exercise or relying on the coach to model everything. In VESSEL, we structure our support the other way around: Support starts as low as possible, and builds up to the level that learners need to proceed. To define when each type of support is given, we must first determine the likely moments and locations in the exercise that learners will need support for. We have used Bloom's [219] taxonomy of keywords and Bayles' [213] overview of online banking critical factors to find all potentially *difficult elements* of the website: All pages and links that a learner can potentially click on, and all complex words and terms on pages that the learner must navigate through to complete the exercise. One proactive support utterance of each support level must exist for each difficult element. One utterance of each level is also needed for each critical waypoint of each exercise. We can then define our fading: For every difficult element, the coach must always start proactive support at support level 1, and increase that level every time the learner needs support again for that same element. Support levels are tracked per difficult element, meaning that a higher support level for one element does not impact other elements. Support levels can only go up, never down.

Transfer of responsibility means that learners must take their own responsibility for the success of the learning process. In VESSEL, this follows automatically from all other scaffolding steps. As learners move through an exercise, proactive support always starts at a low support level and gradually increases, encouraging learners to overcome challenges by themselves instead of waiting for help. Reactive support triggers on learner questions, encouraging learners to actively seek help when needed. And the coach's support delay ensures the gradual lessening of proactive support as learners become more capable of doing everything alone.

5.2.3. Human Factors Knowledge: User Modeling

User modeling refers to the notion of intelligent systems inferring user traits from observable user-system interaction. Fischer [224] defines a user model as "*models that systems have of users that reside inside a computational environment*" (p.70). User models can enable and support advanced user-system interaction by (i.a.) providing user-specific accessibility options [225], limiting the functionality a program provides to match inferred user needs without overloading them [224, 226], and informing users of interaction possibilities and functions that they were not aware of [224, 225, 237]. In the specific context of education and learner support, user models are used to (i.a.) enable adaptive educational and e-learning systems [238, 239], personalize online learning environments [240], and support learners with particular information access and modality needs [241]. Note that not all instances of system adaptation to user behaviour count as or involve user modeling. For instance, VESSEL's cognitive support model (section 5.2.2) already uses user actions to drive its decision making. However, this is more accurately *task* modeling, not user modeling: The system in this instance is only interested in supporting the user with a specific task in a specific moment, not in building a long-term model of that user.

We aim to employ user modeling in VESSEL to improve learning effectiveness. Specifically, we are interested in adapting the aforementioned support delay to the user's overall performance with the exercises. Lehman, Matthews & Person

Table 5.1: VESSEL ECA coach cognitive support categories. Describes exact rules for creating utterances to match each proactive and reactive support level, and includes example utterances used to explain the phrase 'online banking' and the exercise step 'find the page where you change your personal information'.

Support category	**Description**	**Example**
Proactive Support		
Support level 1: Prompt	This utterance asks the user either if they know the meaning of a particular keyword, or if they understand the next exercise step.	"Do you know what 'online banking' means?"
Support level 2: Explanation	This utterance either answers a preceding prompt on the same topic, or answers a direct user question about a particular keyword or exercise step.	"'Online banking' means: doing banking, on your computer."
Support level 3: Hint	This utterance tells the user that their current action or position in the exercise is not correct, and provides oblique direction: The utterance contains one explicit keyword that references the next step the user should take, but doesn't outright say that this is the case.	"You cannot change your address on this page. Can you see where you can change your personal information?"
Support level 4: Instruction	This utterance directly tells the user what action they should take, as an imperative statement. It uses the same keyword as the preceding hint.	"Click on the word: 'personal information'."
Support level 5: Modeling	This utterance offers to demonstrate the right action to the user.	"Shall I show you where you should go?"
Reactive Support		
User utterance: Recognized keyword	The user asks a question that uses a keyword the coach recognizes. The coach provides an 'explanation' support utterance for that keyword.	"Coach, where do I go to do online banking?" "'Online banking' means: doing banking, on your computer."
User utterance: Unrecognized	The user asks a question that does not use any recognized keywords. The coach uses a general reaction utterance to indicate they do not understand.	"Coach, how do I make an account on this website?" "I'm sorry, I cannot help you with this."
User action: Correct	The user moves to the next exercise waypoint correctly. The coach tells the user they have done this.	(*if the user moves to the 'Personal Information' page*) "Well done! The right page for you is 'Personal Information'!"
User action: Incorrect	The user attempts to move to the next exercise when not all correct steps have been taken. The coach tells the user they have made a mistake.	(*if the user fills out the wrong address and then tries to submit their address change*) "Sorry, you have not yet filled out all information correctly."

[242] suggest that struggling learners must be helped along quickly and decisively, which we hypothesize we can do by lowering the delay. Conversely, we hypothesize that increasing the delay for successful learners gives them more time to complete exercises themselves, which will lead to optimal self-efficacy gains by encouraging transfer of responsibility. In both cases, this adaptation should be automatic, or driven by the system, rather than human-invoked [225].

We create a small, simple user model for VESSEL, that encompasses the entire possibility space of all exercises. This is possible because VESSEL forms a relatively compact "closed-world" system [cf. 224], and we can clearly define an optimal path through and an optimal outcome for each exercise. The user model consists of two elements: The user's overall support delay value, and the user's performance in previous exercises. Whenever the user completes a new exercise, the model evaluates their performance in this exercise, and the learner's need for support, by looking at the types and amount of support they needed to pass each critical waypoint in the exercise. If the user passed most waypoints with no support at all, or with prompt or explanation support, their performance in the exercise is rated 'good', and the model increases their support delay by a certain amount. If the user mostly needed instruction and modeling support, their performance is 'bad', and the model decreases their support delay. If the user passed most waypoints with hint support, their performance is 'medium': The balance between challenge and support is right for this user, so their support delay is not changed.

The user model thus outlined serves several purposes. First, using this model, VESSEL can quickly and unobtrusively adapt itself to individual learners. This allows us to present a simple unified VESSEL design at design time, but easily adapt to the needs of users at use time [224, 225]. Second, the model allows VESSEL to reach each user's optimal support delay over time, defined as the support delay in which the user consistently falls in the 'medium' category. As user skill levels improve over time, VESSEL will automatically follow suit. Finally, over longer periods of use, the model would allow us to track users' support delay progress and exercise performance over time, enabling more accurate learning assessment. However, this level of application lies outside the scope of the current work.

5.2.4. Technology: Speech Recognition Rules

In VESSEL, speech recognition is necessary to enable reactive coach support to learner questions (see Table 5.1). The coach can answer questions about the current exercise by recognizing particular *keywords*. We create a dictionary of *known keywords*, which consists of the critical waypoints and difficult elements of each exercise. If the learner says something out loud, the coach checks if any words in the learner's utterance match one of its keywords. If a known keyword is detected, the coach gives explanation-level support about that keyword. If the learner's utterance does not contain any known keywords, it is classified as *unrecognized*. In this case the coach uses a general reaction utterance to indicate lack of understanding, using phrases such as "*I do not understand what you said*". Additionally, the coach can understand the learner utterances "*yes*" and "*no*", allowing it to parse learner answers to questions (see Table 5.1). It can also understand the category of all learner utterances that indicate lack of understanding, such as "*I did not understand that*" and "*Could you repeat what you said*", which ensures that the system is accessible to learners who struggle with quickly interpreting spoken utterances [which includes low-literate second-language learners, cf. 156].

5.3. Specification

5.3.1. Operationalization

In two steps, we translate the updated foundation into comprehensive rules for our ECA coach. First, we formally operationalize the coach's support behaviour during exercises to create the Generalized approach. While the learner works through an exercise, the coach starts a timer that tracks the amount of time that has passed since its last support action. This timer runs continuously regardless of what the learner does, with one exception: The timer is paused whenever learner and coach engage in *learner-coach interaction*, which we define as any dialogue in which both the coach and the learner speak at least once, and the learner's utterances are in reaction to the coach's. Any dialogue that meets these criteria is defined as one occurrence of learner-coach interaction, regardless of length or number of exchanges, with the interaction ending if the learner and the coach do not say anything for five seconds. The timer is temporarily paused while the interaction is ongoing, and resumes when the interaction ends. When the timer exceeds the coach's support delay value, it checks what difficult element the learner is currently interacting with and which critical waypoint the learner should be trying to reach. The coach then gives the proactive support utterance at the support level of that critical element, and resets the timer. If the learner interacts with a difficult element in any way before the support delay value is reached, the coach also resets the timer. If the learner triggers a reactive support utterance (by saying something out loud, or interacting with a waypoint correctly or incorrectly), the coach gives the appropriate utterance and resets the timer. The coach moves through this loop until the exercise is completed. Fig. 5.3 shows the Generalized approach as a decision tree.

Second, we operationalize the Individualized approach, which uses the user model to attune the value of the support delay to learner performance in-between exercises. In this study, we define that the support delay will always be increased or decreased by exactly 5 seconds. The support delay starts at 20 seconds for every learner; it can be raised to a maximum of 30, or lowered to a minimum of 10. See Fig. 5.4 for a visualization of the Individualized approach.

5.3.2. Requirements Baseline

Here we *refine* the existing VESSEL requirements baseline to reflect the updated support behaviour rules; this means we update (expand/re-write) the text of the existing requirements to better reflect our new understanding of the design of VESSEL, and that we write new sub-requirements where necessary. We refine only those requirements that change on the basis of these rules, for the *coach* aspect of VESSEL, the *exercises* aspect, or both. Requirements that are not described in this section stay unchanged. Table D.1 (cf. Appendix D) presents the refined

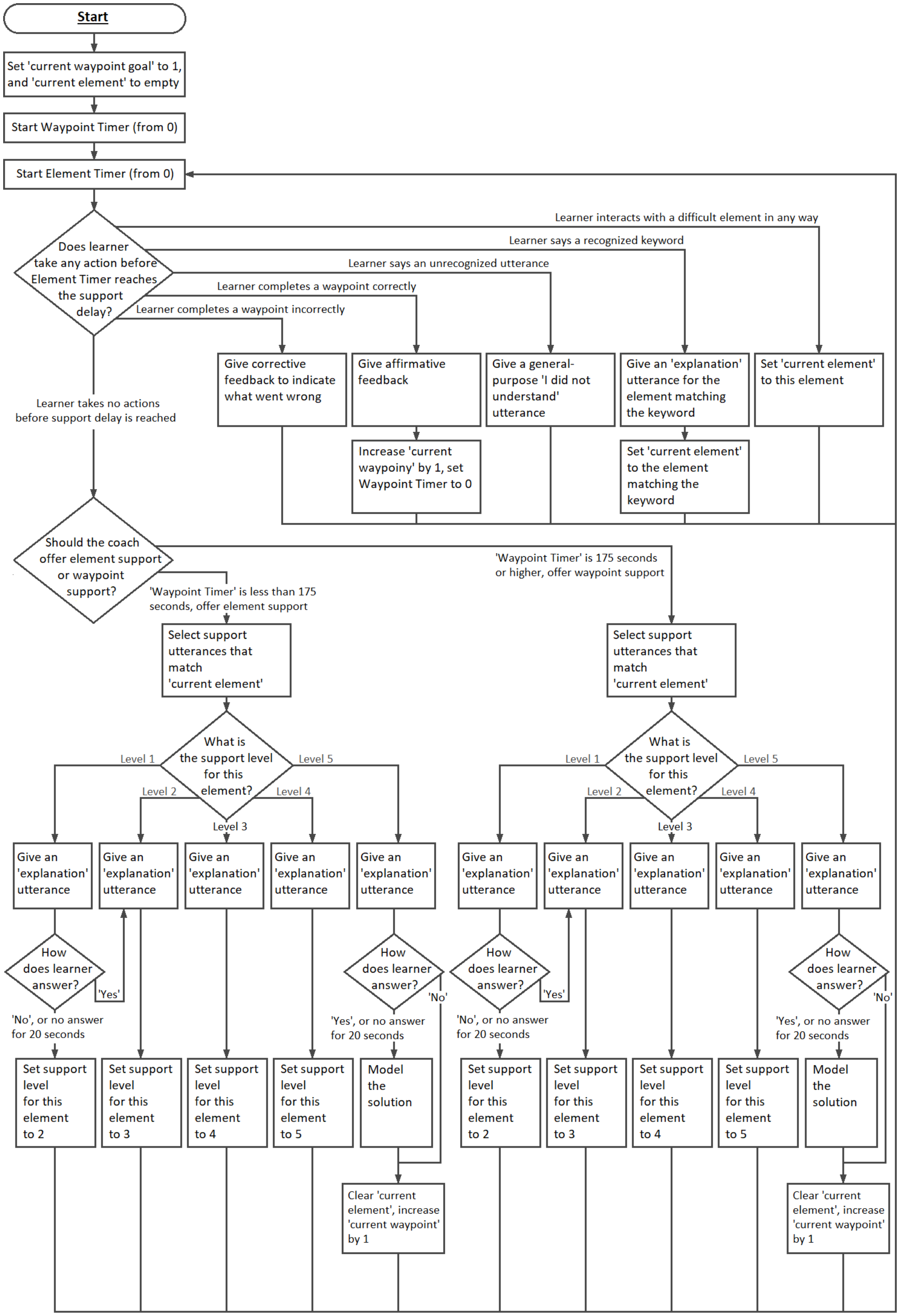


Figure 5.3: Generalized approach rules decision tree. The value of '20 seconds' used here represents the standard support delay.

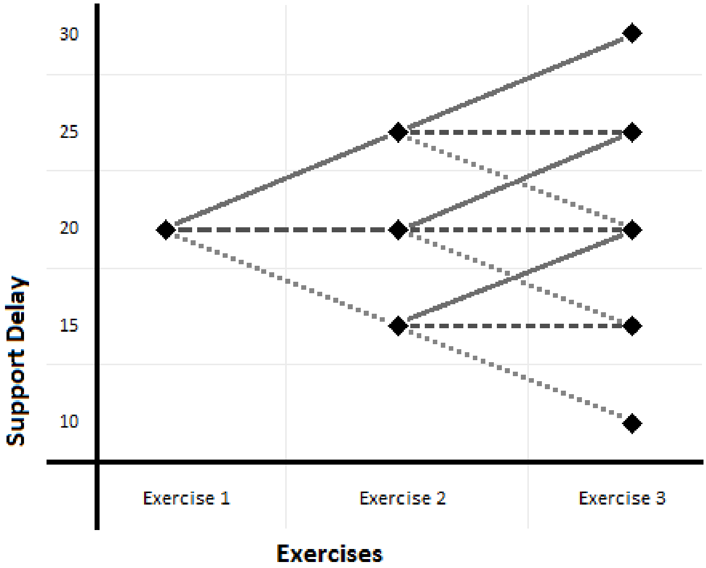


Figure 5.4: Timing schema for the Individualized approach over three exercises. Filled lines represent a learner with 'good' performance, resulting in the support delay being raised, dotted lines represent a learner with 'bad' performance, resulting in the support delay being lowered, and dashed lines represent a learner with 'medium' performance, resulting in the support delay not changing.

requirements baseline.

Requirement **R1. Adaptability** is refined for both the coach and the exercises. The coach should ensure that the support delay best matches the needs of individual learners, using the Individualized approach to attune the delay according to the rules in section 5.3.1 and Fig. 5.4 (**R1.1-C**). And the exercises should be sufficiently challenging to learners. Exercises should exist for different skill and difficulty levels, but these should be neither too easy nor too hard (**R1.1-E**). This can only be evaluated after exercises have been put into practice: An exercise is too easy if learners need little or no coach support to complete it (support on average not exceeding level 1), and it is too hard if learners need strong coach support to complete every step (support on average exceeding level 4). When designing difficulty, it should be kept in mind that the coach's support can lower the difficulty of a too-challenging exercise, but not raise the difficulty of a too-easy one.

Requirement **R6. Support** is zoomed in to only coach-offered cognitive support. The coach should offer cognitive support according to the Generalized approach rules decision tree (Fig. 5.3) (**R6.1-C**).

Requirement **R7. Interactivity** is refined for only the coach. The coach can interact with learners either proactively or reactively. The coach's proactive interaction with the learner should be driven by the support rules decision tree (**R7.1-C**). And the coach's reactive interaction with the learner should be based on section 5.2.4's speech recognition rules (**R7.2-C**).

5.3.3. Use Case: Formalized Cognitive Support for Online Banking

One use case is provided here: The coach giving formalized cognitive support to a learner doing an 'online banking' exercise about transferring money to a different account. Use cases consist of: *Pre-conditions* (conditions that are assumed true at the start of the use case), an *action sequence* (the steps taken by the user and the system over the course of the use case), and *post-conditions* (measurable desired outcomes that result from following the action sequence, i.e. the claims associated with the VESSEL requirements baseline). Two actors are used: '*Coach*' refers to the ECA coach providing formalized cognitive learning support, and '*user*' refers to the low-literate learner engaging with VESSEL. Particular action sequence steps reference Table D.1's (cf. Appendix D) requirements to indicate that this step meets the requirement. Six claims are incorporated: Cognitive/affective/social learning experience, and cognitive/affective/social learning outcomes. Accessibility claims are not used because the user is presumed to already be working with VESSEL.

Pre-conditions:

1. The user is interacting with the coach-supported VESSEL system.
2. An online banking exercise has been selected.
3. The coach and the online banking website are both visible to the user.

Action sequence:

1. The coach introduces the goal and the scope of the exercise to the user. (R.1.1-E, R2.1-C, R3.1-E, R5.1-C)
2. The user uses mouse and keyboard to interact with the online banking website, and a microphone to talk to the coach. (R7.2-C, R7.1-E)
3. Since the coach is using the Individualized approach, it checks the user model for this particular user. Since the user has been successful at previous exercises, the coach sets this user's support delay to 25 seconds. This value will be used throughout the exercise. If the coach had not been using the Individualized approach, it would have set a support delay of 20 seconds without looking at the user model. (R1.1-C)
4. The user tries to navigate to the correct page on the online banking website, but takes a long time doing so. After 25 seconds of the user not making any progress, the coach offers the first level of cognitive support: A prompt. (R6.1-C, R7.1-C)
5. The user still cannot find the right page to navigate to. After another 25 seconds, the coach escalates the level of support to level 2: Explanation. (R6.1-C, R7.1-C)

6. The user reaches the right page and starts filling out information. The user encounters a term they do not understand, and ask the coach about it. The coach finds this keyword in its dictionary, and offers explanation-level support about this keyword immediately. (R6.1-C, R7.2-C)

7. The user fills out some data incorrectly, then tries to move on. The coach notices this, and offers corrective feedback. (R2.1-C, R6.1-C, R7.2-C)

8. The user corrects the mistake and completes the exercise. The coach informs the user that the exercise is over. The coach updates the user model with the results from this exercise. Because the user has performed well, the coach increases the support delay to 30 seconds. In the following exercise, this delay will be used. (R1.1-C)

Post-conditions:

1. The user has actively performed the exercise: The user has done at least one exercise step without the coach modeling the correct solution.

2. The user had a positive experience while doing the online banking exercise: The user's mood has either stayed at the same level of valence, or has increased.

3. The user has interacted with the coach: The user has either asked the coach a question, or answered one of the coach's questions.

4. The user has learned about the online banking steps and can recall this information later.

5. The user's self-efficacy with regard to online banking has increased.

6. The user considers the coach to be friendly and helpful.

5.4. Evaluation: Prototype Development

Functionality. The prototype consists of the three online banking exercises described in section 5.2.1, and an ECA coach that offers cognitive learning support according to the Generalized and Individualized approaches described in sections 5.2.2 and 5.2.3. For the purpose of evaluation, the coach is designed to be controlled via the Wizard-of-Oz method [179].

Interaction methods. Learners interact with the online banking websites using mouse and keyboard. Learners can talk to the coach in natural language. The Wizard operator uses the Fig. 5.3 decision tree to select what utterance the coach says at what moment, choosing pre-recorded spoken utterances from a list. In case of unexpected user actions or utterances, the Wizard can also use the set of general reaction utterances to get the exercise back on track without interruption.

Figure 5.5: VESSEL coach ECA (top right) and summary instructions (in Dutch) for online banking exercise 3.

5

Appearance. The visual appearance of the ECA coach used in Schouten et al. [223] is re-used here. See Fig. 5.5. The coach ECA has one facial animation (opening and closing its mouth while sound is playing, to visually convey that it is 'speaking'), and no gestures or body language.

5.5. Evaluation: Methods

5.5.1. Experimental Design

An experiment was carried out to evaluate the learning effectiveness impact of our formalized-coach VESSEL prototype, as well as to compare the relative effectiveness of the Generalized and Individualized approaches. We therefore used the six learning effectiveness claims that were presented as use-case post-conditions: Cognitive, affective, and social learning experience, and cognitive, affective, and social learning outcomes. Six high-level hypotheses were drafted corresponding to these six claims. Each hypothesis was then zoomed in on two predictions: One prediction about the overall system impact, and one prediction comparing the Generalized and Individualized approaches.

Learning Experience

- **H1. Cognitive Experience (Performance)**
  - **H1a.** The learner takes active part in the exercise: The amount of instruction/modeling support needed to complete exercises is less than 100% of the possible maximum.
  - **H1b.** Learners who receive support along the Generalized and Individualized approaches require less coach support to complete exercises than learners who receive only Generalized-approach support, and expend less subjective mental effort.
- **H2. Affective Experience (Positive Affect)**

- **H2a.** The learner's affective state does not get more negative after completing an exercise with formalized coach support.
- **H2b.** The affective state of learners who receive Generalized and Individualized support changes more positively than learners who receive only Generalized support.

- **H3. Social Experience (Engagement)**
  - **H3a.** The number of learner-coach interactions (*defined in section 5.4*) is more than 0 during an exercise with formalized coach support.
  - **H3b.** Learners who receive Generalized and Individualized support interact with the coach less often than learners who receive only Generalized support.

Learning Outcomes

- **H4. Cognitive Outcomes (Success)**
  - **H4a.** The learner scores more than 0 points on the recall test after completing three exercises with formalized coach support.
  - **H4b.** Learners who receive Generalized and Individualized support take less time to complete any exercise, and score higher on the recall test after completing all three exercises, than learners who receive only Generalized support.
- **H5. Affective Outcomes (Self-Efficacy)**
  - **H5a.** The learner's self-efficacy about online banking increases after completing an exercise with formalized coach support.
  - **H5b.** The self-efficacy increase of learners who receive Generalized and Individualized support is higher than learners who receive only Generalized support.
- **H6. Social Outcomes (Retention)**
  - **H6a.** The learner judges the formalized coach as being helpful and friendly.
  - **H6b.** Learners who receive Generalized and Individualized support judge the coach as more helpful and friendlier than learners who receive only Generalized support.

To test these hypotheses, a mixed-method repeated-measured experiment was designed, combining within-subjects and between-subjects measurements. The study's main independent variable was **Support Model**, with two levels: *Generalized Model* and *Individualized Model*. Participants were invited to complete the three online banking exercises in three experimental sessions, each one week apart: Participants did Exercise 1 in the first week, Exercise 2 in the second week, and Exercise

3 in the third week. Participants were randomly assigned one of two conditions at the start of the first week: 50% of participants worked in the Generalized Model condition throughout the entire experiment, wherein only the Generalized approach was used to provide support, and 50% of participants worked in the Individualized Model condition throughout the entire experiment, which used both Generalized and Individualized approaches.

5.5.2. Measures

Nineteen quantitative dependent variables were measured. Fifteen were self-report questions, measured using three questionnaires (section 5.4), and four were objective performance metrics. Table 5.2 shows an overview of the variables.

5.5.3. Participants

Participants for the study were selected using Kurvers, Dalderop & Stockmann's [89]'s language learner profiles, which subdivide first-language learners (L1) and second-language learners (L2) into five categories. Only learners that matched profiles 2 (fairly skilled L1 and L2 learners), 3 (L2 learners of average skill), and 4 (L1 learners of low skill) were invited to participate, as learners in profiles 1 (highly-skilled L1 and L2 learners) and 5 (L1 and L2 learners with serious learning difficulties) are respectively too skilled to benefit from our level of support, and too low-skilled to engage with the prototype at all. Because the same selection procedure was used in our previous work [223], we also assumed that these participants would have similar information and communication skill levels. Practically, this means we assumed that participant formal information skill levels [information skills in social settings characterized by rigid impersonal rules, such as online banking, cf. 78] were lower than their formal communication skill and informal information/communication skill levels (related to social settings characterized by flexible personalized rules). Participants were recruited from reading and writing classes throughout the Netherlands. Twenty-eight low-literate participants completed the entire experiment: Twenty-one men and seven women, with ages ranging from 24 to 73 (*M=52.1, SD=12.3*). Nineteen of the participants identified as natively fluent in Dutch; the other nine identified as 'somewhat fluent'. Other languages spoken by the participants (either natively or as a second language) included: Arabic, Aramic, Bosnian, Edo, English, French, Hindustani, Italian, Papiamentu, Russian, Somali, Spanish, and Turkish. Eight participants reported prior experience with online banking; of those, seven participants considered online banking easy to do. The twenty participants without online banking experience all found online banking hard.

5.5.4. Materials

The experimental setup consisted of two laptops, each connected to one external monitor (Fig. 5.6), which were used by the experimenters to run the experiment. The external monitors were used by the participants to see and interact with the exercises. The left laptop and monitor were used for the online banking exercises, and the right laptop and monitor were used for the coach. On the participant side,

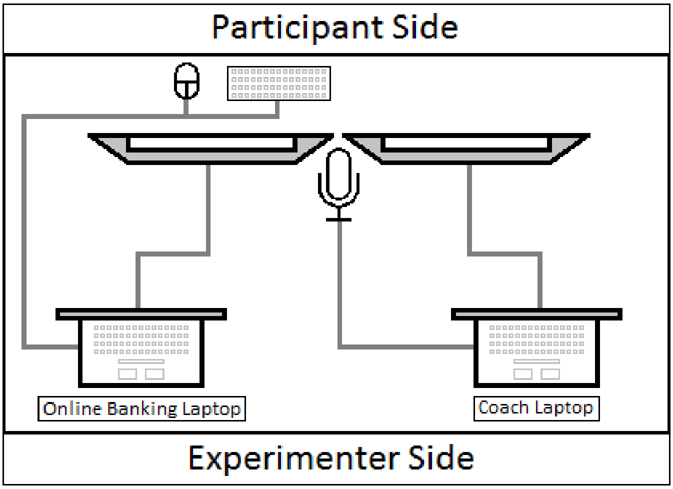


Figure 5.6: Schematic overview of experimental setup. Two monitors (upper figures) are connected to two laptops (lower figures). Keyboard and mouse on participant side are connected to Online Banking Laptop; Microphone placed between monitors is connected to Coach Laptop.

5

a mouse, keyboard, and microphone were provided as well; the microphone was used to 'explain' how participants were able to talk to the coach, as well as to record audio of the proceedings (with consent).

Four questionnaires were used. Three questionnaires measured the fifteen self-report variables (see Table 5.2). First, the 'societal participation questionnaire' (SPQ) measured participant self-efficacy about four example crucial practical situations: Taking out insurance [a representative example of an information skill used in a formal social context, cf. 156], talking at a service desk (communication skill in a formal context), reading a map (information skill in an informal context), and talking to neighbours (communication skill in an informal context). Second, the 'self-assessment questionnaire' (SAQ) measured participant self-efficacy regarding the exercise, and participant affective state. Third, the 'exercise results questionnaire' (ERQ) measured subjective mental effort, and participant affect towards the coach. Two answer methods were used: A visual analogue scale (Fig. 5.7), and the Self-Assessment Manikin (SAM) 5.8. Answers to self-efficacy, mental effort, and coach affect questions were given using the visual analogue scale, as this method does not require reading and writing skills, and allows participants to rate concepts that are otherwise hard to describe or categorize [243]. Answers to self-affect questions were given using the SAM, which measures three affective dimensions: Pleasure/valence, arousal, and dominance [244]. Questions were always read aloud to participants, who would then mark their answer on the matching bar or figure. The fourth 'demographic' questionnaire measured participant age, sex, schooling history, time spent in the Netherlands, languages known, and prior experience with online banking. These questions were read out loud as well; the researchers wrote down the answers.

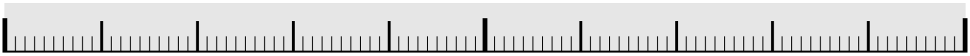


Figure 5.7: Visual analogue scale used to measure self-efficacy, subjective mental effort, and coach affect.

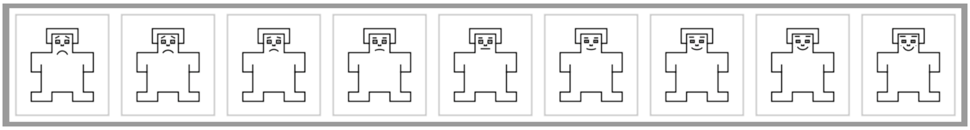


Figure 5.8: Self-Assessment Manikin used to measure participant pleasure/valence.

In addition to the questionnaires, four objective measures were taken. First, participant completion time was measured with a stopwatch. Second, exercise support level was calculated by tabulating the number of times each coach utterance type (Table 5.1) was used in an exercise and dividing the sum of the resulting support levels (1 for prompts, 2 for explanations, et cetera) by the number of critical waypoints. Third, learner-coach interaction was recorded with the microphone. Lastly, a 'recall test' was created to measure participants' learning success. The test consisted of six A4-printed screenshots of the online banking website. For each of the six pictures, participants were given 60 seconds to answer one question, referencing an activity from one of three exercises. Answers were scored as either fully correct (1 point), partially correct (.5 points), or incorrect/out of time (0 points).

5.5.5. Procedure

The three experimental sessions were held over the course of three weeks, each one week apart. Two researchers were present: One researcher acted as the dedicated Wizard-of-Oz controller for the coach, while the other managed all participant interaction and controlled the online banking task environment. The first session started with general introduction, informed consent forms, and the demographic questionnaire. The first SPQ was administered, followed by the first SAQ. The managing researcher explained the general experiment flow, and activated the coach, which was controlled by the second experimenter. The coach introduced itself to the user, explained the first exercise, and showed the instruction material. Participants were told to complete the first exercise with the help of the coach. No time limit was set. As soon as participants were finished, researchers administered an ERQ and a second SAQ. Participants were then debriefed, ending the first session. In-between the first and second session, all participants' performances were rated, using the 'good/medium/bad' categorization described in section 5.2.3. For participants in the Individualized condition, the user model was updated and support delays were changed where necessary (as shown in Fig. 5.4).

In the second session, researchers started by administering an SAQ. After that, flow proceeded as per the first session, with participants completing the second exercise before filling out an ERQ and an SAQ. In-between the second and third session, participant performances were again rated, and support delays were again

Table 5.2: Overview of measures. Includes measure source (societal participation questionnaire, self-assessment questionnaire, exercise results questionnaire, or direct measurement) and description.

Variable	Description
Subjective measures: societal participation questionnaire (SPQ)	
SPQ.1. Self-efficacy (formal information skill)	”I can take out insurance.”
SPQ.2. Self-efficacy (formal communication skill)	”I can ask for help at a service desk.”
SPQ.3. Self-efficacy (informal information skill)	”I can read a map.”
SPQ.4. Self-efficacy (informal communication skill)	”I can talk to my neighbours.”
Subjective measures: Self-assessment questionnaire (SAQ)	
SAQ.1. Self-efficacy (reading Dutch)	”I can read Dutch.”
SAQ.2. Self-efficacy (online banking)	”I can do online banking.”
SAQ.3. Self-efficacy (computer use)	”I can use a computer.”
SAQ.4. Affect (valence)	”How good do you feel right now?”
SAQ.5. Affect (arousal)	”How active do you feel right now?”
SAQ.6. Affect (dominance)	”How strong do you feel right now?”
Subjective measures: Exercise results questionnaire (ERQ)	
ERQ.1. Subjective mental effort	”How much effort did it take you to complete the exercise?”
ERQ.2. Coach-affect (valence)	”The coach was happy.”
ERQ.3. Coach-affect (arousal)	”The coach was busy.”
ERQ.4. Coach-affect (dominance)	”The coach took charge.”
ERQ.5. Coach-affect (usefulness)	”The coach helped with the exercise.”
Objective measures: Direct measurement per exercise	
DM1. Completion time (seconds)	Time from start of exercise to completion.
DM2. Level of coach support	Highest level of coach support needed to pass any waypoint.
DM3. Learner-coach interaction	Amount of learner-coach-interaction during the exercise.
DM4. Recall test score	Score on end-of-experiment recall test.

5

updated for participants in the Individualized condition. The third session (with the third exercise) was similar to the previous two, except for additions at the end: After the final exercise results and SAQ, researchers administered a second SPQ. After this, the recall test was explained and administered. Finally, participants were fully debriefed (including a 'look behind the scenes' for the Wizard-of-Oz method, and a short qualitative interview to see how they experienced working with the prototype and the coach) and rewarded for participation.

5.6. Evaluation: Results

Three analysis steps were done. First, the data were characterized and starting assumptions were checked, by looking at participant descriptives, exercise difficulty levels, and the effectiveness of the different support levels. Second, quantitative analyses were conducted on the Table 5.2 measures in order to verify the hypotheses. And third, two post-hoc analyses were carried out: The predictive value of several variables on recall test score was tested, and groups of participants were evaluated based on initial performance. Finally, qualitative observations were made by the researchers, during the experiment and by listening to the audio recordings afterwards.

Before analysis, data validity was checked in four ways, following Nimon's [245] outline of statistical assumptions in General Linear Model (GLM) analyses. First, P-P and Q-Q plots were used to assess multivariate normality. Results showed that multivariate normality was upheld for all measures except three: Measures SPQ.3 and SPQ.4 show mild and medium abnormality, respectively. And while measure DM2 shows a good normal distribution, dividing this variable into DM2a and DM2b (see also Table 5.4) shows that while DM2a is normally distributed, DM2b is mildly abnormal. Second, Mauchly's test of sphericity was used to assess data variance. Results showed that the assumption of equal pair variance was upheld for all measures except measure SAQ.1. Third, questionnaire reliability was assessed. Cronbach's α was .730 for the SPQ, .872 for the SAQ, and .734 for the ERQ. No data reduction measures were used. Fourth, the dataset was checked for overall correctness. Logging issues were discovered in the support level data for three participants; these participants were excluded from further support level analyses (pertaining to DM.2 and DM.3), but otherwise included. Given these results, we were confident to proceed with the planned analyses.

5.6.1. Assumptions

Four assumptions were checked: The assumption of participant starting skill, the assumption of equal exercise difficulty, the assumption of support model effectiveness, and the assumption of temporal contingency. The *assumption of participant starting skill* was that the formal information skill level for low-literate participants would be low when compared to their formal communication skill and informal information/communication skills. The *assumption of equal exercise difficulty* was that all three exercises would require similar amounts of time and support to complete. The *assumption of support model effectiveness* was that, from prompt to modeling,

Table 5.3: Societal participation questionnaire means and standard deviations.

	Pre-Experiment	**Post-Experiment**
SPQ.1. 'I can take out insurance.' (*formal information skill*)	49.86 (*SD=36.98*)	44.82 (*SD=32.18*)
SPQ.2. 'I can get help at a service desk.' (*formal communication skill*)	80.61 (*SD=23.39*)	78.00 (*SD=22.15*)
SPQ.3. 'I can read a map.' (*informal information skill*)	69.43 (*SD=33.50*)	64.39 (*SD=32.38*)
SPQ.4. 'I can talk to my neighbours.' (*informal communication skill*)	86.86 (*SD=21.79*)	81.50 (*SD=25.47*)

the five utterances in the support model would be more effective at helping learners complete exercise steps. The *assumption of temporal contingency* was that a coach with a lower support delay (with 10 seconds being the lowest possible delay and 30 seconds the highest) would result in a higher average support level, and a lower average exercise completion time.

To check the assumption of participant starting skill, SPQ means were compared with a paired-samples T-test (Table 5.3). Analysis shows that before the start of the experiment, participants rated their formal information skill (SPQ.1) as significantly lower than their formal communication skill (SPQ.2, $t(27)=-4.313$, $p=.000$), informal information skill (SPQ.3, $t(27)=-2.657$, $p=.013$), and informal communication skill (SPQ.4, $t(27)=-5.413$, $p=.000$). Informal information skill was also rated as lower than informal communication skill ($t(27)=-3.049$, $p=.005$). Post-experiment, the exact same pattern was seen (respectively ($t(27)=-5.396$, $p=.000$), ($t(27)=-2.918$, $p=.007$), ($t(27)=-5.670$, $p=.000$), and ($t(27)=-3.228$, $p=.003$)). As such, this assumption was upheld.

To check the assumption of equal exercise difficulty, a repeated-measures GLM analysis compared exercise completion time and average support level for the full exercise, as well as support level for only the navigation steps and support level for only the data entry steps. Table 5.4 shows the results of the analysis. Significant differences were found: The second exercise required a lower overall support level to be completed, the third exercise required a lower navigation support level, and all three exercises required different amounts of data entry support. As such, the assumption of equal difficulty was not upheld. In light of these findings, we chose not to alter our *a priori* planned hypotheses evaluations, but to incorporate these findings into a post-hoc analysis (section 5.6.3).

To check the assumption of support model effectiveness, we tabulated the total number of support utterances given for each level. We also counted how many utterances in each level successfully helped a participant get to the next critical waypoint; i.e. if the instruction 'click on the word Online Banking' got a participant to navigate to the online banking page, then that utterance was successful. Table

Table 5.4: Exercise descriptives. Completion time is measured in seconds. 'Average support level' means: The average highest level of support needed to pass critical waypoints. F-value (F), significance (p), and observed power (β) are given if p<0.05.

	1st Exercise	**2nd Exercise**	**3rd Exercise**	**Test statistic**
DM1. Average completion time (in seconds)	691 (*SD=302*)	568 (*SD=232*)	704 (*SD=315*)	
DM2. Average support level (all waypoints)	2.02 (*SD=1.06*)	1.58 (*SD=1.03*)	2.03 (*SD=1.25*)	F(2,23)=5.183 p=.014 β=.774
DM2a. Average support level (navigation waypoints)	2.74 (*SD=1.04*)	2.71 (*SD=1.30*)	1.79 (*SD=1.37*)	F(2,23)=9.117 p=.001 β=.956
DM2b. Average support level (data entry waypoints)	1.29 (*SD=1.33*)	0.44 (*SD=1.07*)	2.27 (*SD=1.36*)	F(2,23)=26.245 p=.000 β=1.000

Table 5.5: Number of utterances given for each support type, and success rate for each, over the entire experiment. 1124 support utterances were recorded in total.

	Prompt	**Explanation**	**Hint**	**Instruction**	**Modeling**
Number given	329	290	253	166	85
Number successful	38	38	87	81	85
Success rate	11.6%	13.1%	34.4%	48.8%	100%

5.5 shows the number of utterances for each category, as well as the success rate. The numbers show that in the order of prompt, explanation, hint, instruction, and modeling, the success rate of each utterance goes up. As such, this assumption was upheld.

Finally, to check the assumption of temporal contingency, one-way ANOVA analyses were done on the average support level and average completion times of exercises 2 and 3, using coach support delay for that exercise as an input. Exercise 1 was not used, as all participants had a support delay of 20 seconds in that exercise. Table 5.6 shows that as the coach's support delay went down, the average support level increased (exercise 2: F(2,23)=5.755, p=.010; exercise 3: F(3,22)=4.555, p=.013), but average completion time did not decrease as expected. We chose to continue with our envisioned hypothesis evaluations, and to keep these findings on hand when interpreting the results of any analysis that leans on the assumption of temporal contingency.

Table 5.6: Average support level and completion times for exercises 2 and 3, per coach support delay category. Rows marked with * show significant ANOVA differences at p<.05. Columns marked 'X' are not relevant: in exercise 2, 10s and 30 timings were impossible to reach by design.

	10s	**15s**	**20s**	**25s**	**30s**
Exercise 2					
Number	X	4	14	7	X
Average support level*	X	2.65 (*SD=1.24*)	1.65 (*SD=.91*)	.82 (*SD=.47*)	X
Average completion time (s)	X	654 (*SD=215*)	566 (*SD=241*)	396 (*SD=204*)	X
Exercise 3					
Number	2	0	13	4	6
Average support level*	3.81 (*SD=.97*)	-	2.24 (*SD=1.19*)	2.16 (*SD=.53*)	.90 (*SD=.86*)
Average completion time (s)	961 (*SD=238*)	-	732 (*SD=295*)	838 (*SD=235*)	490 (*SD=217*)

Table 5.7: Hypothesis evaluation table. Legend: B1/A1 means 'Before exercise 1'/'After exercise 1', etc. Column 'Hypothesis Hxa' contains test statistics that evaluate hypotheses H1a to H6a. 'Rep. GLM' means repeated measures GLM. '1s-T (X)' means one-sample T-test, testing against H0=X. '1-way ANOVA' means one-way ANOVA on experimental condition. Test results are given for p<.05. Grey boxes mean no value was measured or no test was conducted, blank boxes mean no significant result was found.

5.6.2. Hypotheses Evaluation

To evaluate hypotheses H1 through H6, the data from the SAQ, ERQ, and the direct measurements (see Table 5.2) were systematically analyzed. Table 5.7 shows a schematic overview of all data measurements, ordered per hypothesis. Included in the table are: Means and standard deviations per measurement moment (before/after exercise 1/2/3), which statistical test was used to analyze the measure, and the relevant test statistic, if significant. Three types of tests were used: Repeated measures General Linear Model analysis, one-sided T-tests, and one-way ANOVAs. All GLM analyses were done using all data points as one factor (meaning they all contained either one factor with three levels, or one factor with six levels); additionally, all GLM analyses were conducted either without any between-subjects factors (for hypotheses H1a to H6a), or using participant experimental condition as a between-subjects factor (for hypotheses H1b to H6b). One-way ANOVAs were conducted on participant experimental condition. One-sided T-tests were conducted on select values, as shown in Table 5.7. Note that Table 5.7 only shows hypothesis evaluations for H1a to H6a; evaluation of hypotheses H1b to H6b showed no significant results, and as such was not included in the table.

The following results were found. For all exercises, the average support level

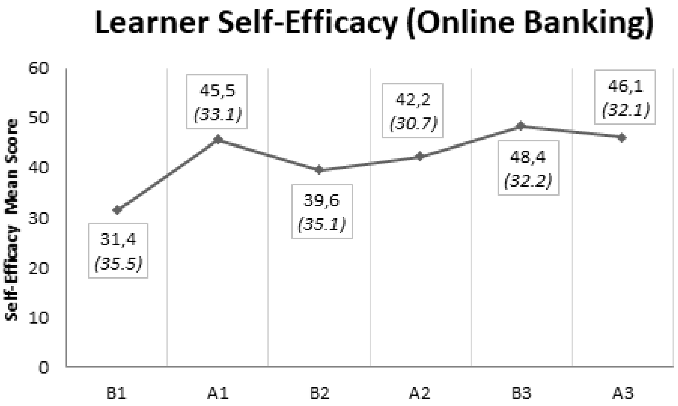


Figure 5.9: Mean of 'online banking self-efficacy' for the six measurement moments. Boxes indicate: mean (standard deviation). Horizontal axis shows the six measurement moments. Vertical axis shows score on SAQ.2, in range [0-100], measured using the visual analogue scale (Fig. 5.7). Columns 'B1' to 'A3' refer to measurement moments 'Before exercise 1' to 'After exercise 3'.

5

was lower than 4, indicating no exercise required instruction and/or modeling support for every critical waypoint. This supports H1a. For measures SAQ.4, SAQ.5, and SAQ.6, repeated-measures GLM shows no significant differences across exercises. This indicates participant affective state did not get significantly more negative as a result of working with the coach, supporting H2a. On average, participants interacted with the coach more than 0 times in each exercise, supporting H3a. On average, all participants scored higher than 0 on the recall test, supporting H4a. A closer look at the data shows that no single participant scored 0 on the test. Measure SAQ.2 (Self-efficacy – online banking) was significantly different across exercises. Follow-up analysis shows that value B1 ('before exercise 1') was significantly lower than the other five, indicating that online banking self-efficacy has increased after completing exercise 1. Fig. 5.9 shows this result. This partially supports H5a, as self-efficacy does not increase after every exercise. Finally, one-sided T-tests show that the averages of ERQ.2, ERQ.4 and ERQ.5 are significantly higher than the scale midpoint, and that ERQ.3 is significantly lower. This suggests that participants judged the coach as affectively positive, calm, dominant, and helpful, weakly supporting H6a.

Finally, tests for between-subjects effects showed no significant results for age, sex, schooling history, time spent in the Netherlands, languages known, and prior experience with online banking.

5.6.3. Post-Hoc Analyses

Recall Test Analysis

Regression analyses were carried out to test if the following variables could predict recall test scores: Average support level throughout all exercises, completion time per exercise, average completion time across all exercises, participant age, participant sex, participant experience with online banking, and number of weeks spent living in the Netherlands. Prior to this, a bivariate correlation analysis was carried out to see which variables should be included in a single regression test. This analysis showed that several variables were significantly correlated (at $p<.05$), limiting their applicability for regression analysis. The following variables were selected for a stepwise linear regression for knowledge test results: Average support level, participant sex, time spent in the Netherlands, and experience with online banking. One significant result was found: Average support level negatively predicts knowledge test results ($t=-3.806$, $p=.001$). A curve estimation analysis was done to confirm this. Linear, quadratic, and logarithmic models were tested. Both a linear model ($F(1,23)=14.483$, $p=.001$) and a logarithmic model ($F(1,23)=19.708$, $p=.001$) confirmed that a higher average support level corresponded to a lower recall test score. See Fig. 5.10.

Performance Group Analysis

One interpretation of the preceding hypothesis and recall test analyses is that participant online banking skill levels did not change significantly over the course of three exercises. In this case, 'participant online banking skill level' should be treated as a set trait instead of a dependent variable. If all three exercises were equal in challenge, exploratory techniques (e.g. cluster analysis) could reveal this. However, Table 5.4 shows that the exercises are not equal in terms of the level of support needed to complete them: Website- and exercise-specific learning effects in the 2^{nd} and 3^{rd} exercises conflate the grouping. This strongly implies that some exercises were more challenging or difficult than others. Taking this into account, we clustered participants into three 'performance groups' based on their performance in the 1^{st} exercise; we made the assumption here that their performance in this exercise was the most accurate reflection of their 'actual' online banking skill level, before any potential learning effects from the experiment and the effects of individualized support came into play. Six people were assigned to the 'Bad' group, ten people to the 'Medium' group, and nine people to the 'Good' group, based on their performance in the first exercise, using the established user model categorization (section 5.2.3). The repeated-measures GLM analyses in section 5.6.2 were then run again with this variable as a between-subjects factor with three levels: *Bad*, *Medium*, and *Good*. Two effects were found: Compared to Medium and Good, participants in the Bad group had significantly lower computer-use self-efficacy overall (main effect of between-subjects factor, $F(2,23)=5.402$, $p=.012$, Fig. 5.11), and (on average) dropped in positive affect after completing any exercise (interaction effect, $F(2,23)=3.525$, $p=.047$, Fig. 5.12). As a result of this last finding, hypothesis H2a is no longer fully supported, but partially supported: The affective state of participants in the Good and Medium groups did not get worse as a result of working with the coach, but the affective state of participants in the Bad group did.

5.6.4. Observations & Interviews

Experimenters observed that participants managed to work with the coach as intended. The provided support was sufficient for the exercises: Almost all partici-

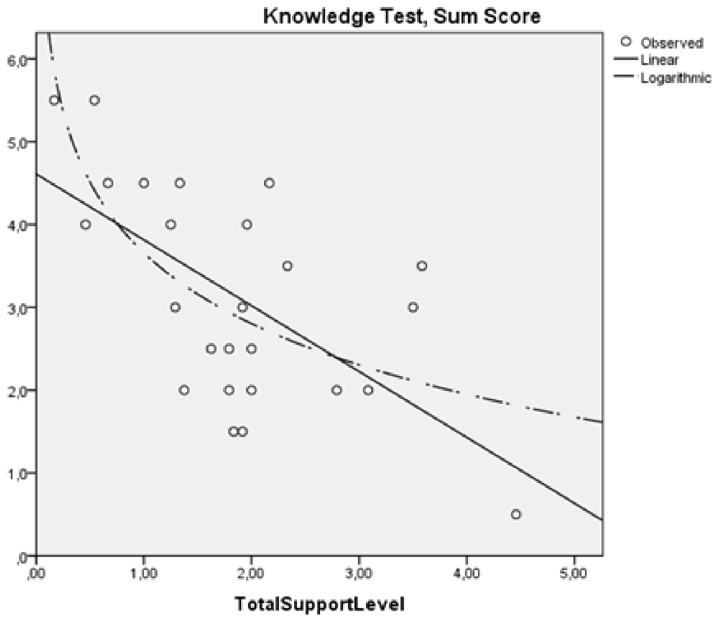


Figure 5.10: Curve estimation result for recall test score as a function of average support level throughout exercise.

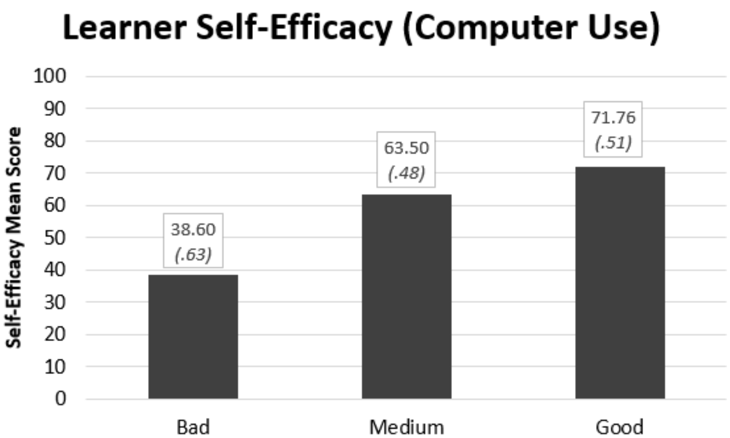


Figure 5.11: Performance group analysis showing main between-subjects effect on SAQ.3 (computer use self-efficacy). Boxes indicate: mean (standard deviation). Horizontal axis shows the three performance groups. Vertical axis shows score on SAQ.3, in range [0-100], measured using the visual analogue scale (Fig. 5.7).

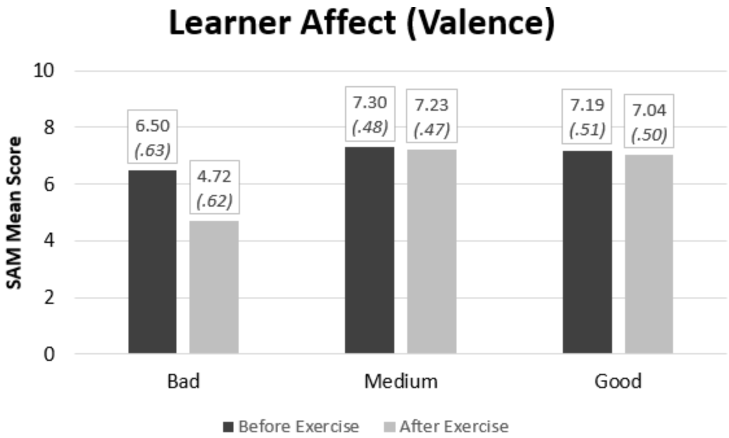


Figure 5.12: Performance group analysis showing interaction effect on SAQ.4 (valence). Boxes indicate: mean (standard deviation). Horizontal axis shows the three performance groups; two bars per group indicate measures taken before and after any exercise. Vertical axis shows score on SAM.1, in range [1-9], measured using the Self-Assessment Manikin (Fig. 5.8).

pants took active part in the exercises, even when these were obviously difficult, and managed to complete them fully. Only three times did participants 'give up' and wait for the coach to model every remaining step. While doing the exercises, participants listened to the coach's support and generally followed direct instructions if they understood them. Participants successfully interacted with the coach within the constraints of our speech recognition and support behaviour rules. The experimenters felt that the 20-second support delays were very long, and that for particular participants (e.g. participants who would switch their attention around very quickly) the support utterances did not arrive 'at the right time'. But on the participant side, this was not experienced. In interviews, participants simply accepted that the coach was slow sometimes, and that 'she took some time to give a good answer'.

Different progress results were seen in the different support delay timing conditions. In the 20-second condition, most participants were able to complete all critical waypoints without requiring instruction or modeling support. And support was often given while participants were actively engaged with the exercise. Similar patterns were seen in the 15- and 25-second conditions (but this is limited by the low number of observations in these conditions). Different patterns were seen in the 10- and 30-second conditions. In the 10-second condition, participants received support so quickly that they often had no time to process it before another utterance was due. Many more instances of instruction and modeling support were seen here than in other conditions (also shown in Table 5.6). In the 30-second condition, although many participants in this condition hardly needed help, it was observed that when participants did need help to proceed, they had to sit through long and noticeable waiting times. More so than in other conditions, participants seemed annoyed that the coach would not immediately answer their questions.

Two additional observations stand out. First, while participants did often interact with the coach, experimenters felt as though the total amount of human-coach interaction in this study was lower than in the previous one [223]. Participants that spoke to the coach talked as they would to a human conversation partner, using complete sentences and sometimes even gesturing at the screen. But not all participants spoke to the coach often, or at all. Particularly, while participants often reacted to coach questions and prompting, very little proactive interaction was seen. In Schouten et al. [223], participants very often talked to the coach extensively and in great detail, including asking it highly complex questions and even telling it stories about their own lives. This rarely happened in our current study (although it did happen, with one participant even joking he'd "*like to take [the coach] on a date sometimes*"). And second, while all participants completed all exercises with the coach's help, ending interviews revealed that many reflected on this negatively. Participants did not see the situation as them working together with the coach for the goal of learning, but as them failing to complete a challenge and the coach needing to 'rescue' them. One participant, who completed exercise 3 in good time but needed modeling to find the very last navigation waypoint, complained that "*...I wouldn't have been able to do it without the coach*".

5.7. Conclusions and Discussion

5.7.1. Conclusions

This study intended to answer two research questions. Question Q1 was: "*How can we create a design specification for VESSEL that incorporates rules for cognitive learning support provided by an ECA coach?*" Sub-question Q1a, "*Which operational demands, human factors knowledge, and technologies are needed to write these rules?*", was answered in sections 5.2 and 5.3. In section 5.2, we showed how hard online banking exercises provide a task environment for cognitive learning support, how the scaffolding concepts of contingency, fading, and transfer of responsibility inform the coach's cognitive support behaviour, how user modeling can be employed to adapt offered support to individual performance and circumstances, and how we envision the role of speech recognition. By incorporating this into the foundation, we resolved our knowledge gaps. In section 5.3, we created dialogue rules to specify the ECA coach's cognitive support behaviour, refined the requirements baseline to incorporate these rules, and wrote a new use case to illustrate the envisioned user-system interaction. Sub-question Q1b, "*Which functionalities, interaction methods, and appearances should the ECA coach have to reflect this specification?*", was answered in sections 5.3 and 5.4: A new VESSEL prototype was created on the basis of our specification, including formalized cognitive learning support rules and user modeling functionality.

Question Q2 was: "*What is the learning effectiveness impact of a VESSEL prototype that offers cognitive learning support according to the formal specification?*" Sub-questions Q2a, "*Are the learning effectiveness results of this prototype comparable to the VESSEL prototype that offered informal cognitive, affective, and social learning support?*", and Q2b, "*Does using both the Generalized and Individualized* *approaches to learning support result in higher learning effectiveness than using only the Generalized approach?*", were answered by experimentally evaluating the prototype in sections 5.5 and 5.6. We tested six hypotheses for each sub-question (twelve in total). For question Q2a, hypotheses H1a, H3a, and H4a were fully supported, and H2a, H5a and H6a were partially supported, showing that the ECA coach's formalized cognitive support resulted in high learning effectiveness for low-literate learners. Cognitively, learners used the coach for guidance, but did not rely on it for everything. Affectively, the coach had no negative influence on the user's mood for users in the Good and Medium performance groups. Self-efficacy regarding online banking increased after doing the first exercise, and stayed at the new high level afterwards. And socially, learners interacted with the coach as if it was human, and judged 'her' as a friendly, useful helper. These results suggest that the formalized coach meets our design goals. Learners can use the coach to successfully complete challenging exercises, resulting in non-zero recall and a significant increase in self-efficacy. The lowered affective state of users in the Bad performance group is unexpected, however, and this should be further investigated in future work.

Comparing these results to Schouten et al. [223] shows interesting similarities. Both studies show a significant increase in self-efficacy after completing one coach-supported hard online banking exercise. In both cases, the actual reported values for self-efficacy are just below or around the scale midpoint (0 in Schouten et al. [223], 50 in this study), suggesting that while online banking self-efficacy does *increase*, it is still not very *high*. Other value similarities include positive affect when the coach was present (halfway between scale midpoint and maximum value), difficulty and required effort of the exercise (idem), and the degree to which the coach was seen as a supportive agent (close to scale maximum). These similarities suggest that the learning effectiveness results of this prototype and the Schouten et al. [223] prototype are comparable, answering question Q2a. Importantly, these results seem to indicate that moving the coach to keyword-based speech recognition was not a significant problem for low-literates. Experimenter observations corroborate that low-literate participants had little problems using the coach. Almost all participants asked their questions slowly and clearly, using the exact terms from the website even without being instructed to do so. When problems did occur, it was often because participants used unanticipated question phrasings and keywords, or because they assumed too much real human conversation ability on the coach's behalf: For instance, certain participants expected the coach to be able to use past conversation context, attempting to reference things that happened earlier in the experiment or even in earlier experimental sessions. Further development, including expanded keyword lists and more dialogue rules, could alleviate these problems. Experimenters did feel that there was less learner-started social interaction in this study than in the Schouten et al. [223] study, which both experimenters were also part of. We suspect that this happened because the previous study's coach used small talk for social support at the start of exercises. By asking the learner questions about their life and talking about 'her own experiences', this coach afforded being spoken to like a human partner, acclimatizing low-literate learners to the idea they could actively ask questions. Since the current coach did not do this, participants may not have considered to try. Learners still reactively answered the coach's questions, but would only sometimes proactively ask questions. Future work should study whether small talk influences learner-coach interactions in this way.

For question Q2b, hypotheses H1b through H6b were all not supported. This shows that including the Individualized support model did not significantly improve on the Generalized model. Observed qualitative differences were not reflected in quantitative measures. Two possible explanations can be offered. One (unlikely) option is that support delay does not have a significant influence on the learning experience of low-literates at all. We instead suspect that our manipulations did not actually match learning support to user skills, meaning we did not achieve fading the way we envisioned. Qualitative and quantitative data support this explanation: Lower support delays caused information overload, while higher support delays caused a need for waiting. Future work should verify this: Perhaps smaller delay changes with less extreme end points, over a longer period of time, would result in fading as expected.

5.7.2. Limitations

Three limitations are identified in this study. First, the number of participants recruited for this study is relatively low for the purposes of quantitative statistics. This problem is difficult to avoid when doing experimental research with low-literates, as the available pool of potential participants is low: Finding and recruiting low-literates is a non-trivial issue [cf. 156], and we further limited this pool by using Kurvers, Dalderop & Stockmann's [89] language learner profiles as a selection criterion (see section 5.5.3). This calls the power of our results into question. While analysis has shown that our data upholds multivariate normality and equal pair variance, and observed power was generally satisfactory, a larger sample size would solidify our findings. A stand-out point is the fact that 8 of our 28 participants reported prior experience with online banking, which seems like a strong potential confound. As the online banking environment used in our work was created for this experiment (meaning no participants could have direct experience with it), and as between-subjects analysis showed no significant effect for 'prior online banking experience', we are confident about the accuracy of our findings; nevertheless, future work should give this factor strong consideration.

Second, the experimenters ran into some implementation issues with the prototype. The Wizard operator could not correctly control the coach 100% of the time: Technical difficulties in the coach's control program caused unavoidable time delays of up to 12 seconds between selecting a coach utterance, and that utterance actually playing. This problem was first encountered during pilot testing, but was not resolved before the actual experiments took place. As a result, the Wizard operator had to train to factor them in. This formed a source of noise in the support provision. Additionally, in situations where the participants performed actions the coach was not built to expect, the Wizard had to append new rules on the fly. For instance, at the start of the first experiment, there was no rule for what to do if the participant returned to an earlier-completed waypoint. This situation was encountered during the first experiment, at which point a rule was created to handle it. Afterwards, this rule was incorporated into the coach and executed consistently. However, initial occurrences of situations like this still introduced noise.

Finally, the post-hoc analysis of learner skill and performance reveals a problem with the assumptions underlying our work. Section 5.6.3 shows that learners who required more support to complete exercises scored more poorly on the recall test. One interpretation is that, in our three weeks of testing, learners' actual skill in doing online banking has not changed. Rather, learners with initial high skill levels needed little support and scored well, while learners with low skill levels needed much support and scored poorly. This is supported by the performance group analysis, which shows that learners who performed poorly in the first exercise consistently had a worse mood after exercises, and judged their computer use self-efficacy to be low (section 5.6.3). It looks as if the coach has only helped learners *complete* the exercise, not *master* it. In the interviews (section 5.6.4), participants blamed themselves for failing to succeed alone and viewed the coach's instructions as an admittance of that failure. Two assumptions underlying requirements R1.1-C (coach adaptability) and R1.1-E (exercise adaptability) are that learning support can lower the experienced difficulty of challenging exercises, and that low-literate learners should not be allowed to fail. But results suggest that participants still experienced the exercise as very challenging; the coach's help was not seen as lower difficulty, but as unfair help. Even though all exercises were completed, learners attributed failure on the level of separate critical waypoints to themselves. Future study should investigate if this happens consistently. If the coach cannot actually lower experienced difficulty, and if low-literate learners weigh failure on any waypoint level more heavily than success on the overall exercise, the assumptions underlying our coach's behaviour must be rethought.

5.7.3. Future Work

This study has demonstrated the value of using formalized cognitive learning support for low-literates. Learners successfully interacted with the coach to complete challenging exercises, which resulted in a positive learning experience and higher online banking self-efficacy. These findings indicate that our current VESSEL development direction has merit. We will build on this in future work: Now that our cognitive support has been formalized and evaluated, we can try to do the same for affective and social support. In a next SCE iteration, we turn our attention towards building a prototype that provides support not only contingent on the learner's cognitive needs, but also their affective and social needs [cf. 183].

6

Formalized Affective & Social Learning Support

In this study, we investigate if a digital coach for low-literate learners that provides cognitive learning support based on scaffolding can be improved by adding affective learning support based on motivational interviewing, and social learning support based on small talk. Several knowledge gaps are identified: motivational interviewing and small talk must be translated to control rules for this coach, a formal model of participant emotional states is needed to allow the coach to parse the learner's emotional state, and various sensors must be used to let the coach detect and act on this state. We use the Socio-Cognitive Engineering (SCE) method to update an existing foundation of knowledge with emotional models, motivational interviewing and small talk theory, technology, and a new exercise in the volunteer work domain. We use this foundation to create a design specification for an Embodied Conversational Agent (ECA) coach that provides cognitive, affective, and social learning support for this exercise. A prototype is created, and compared to a prototype that only provides cognitive support in a within- and between-subjects experiment. Results show that both prototypes work as expected: learners interact with the coach and complete all exercises. Almost no significant differences are found between the two prototypes, indicating that the affective and social support were not effective as designed. Potential improvements are provided for future work. Results also show significant differences between two subgroups of low-literate participants, and between men and women, reinforcing the importance of using individualized support measures with this demographic.

The content of this chapter has been published as: Schouten, D. G. M., Venneker, F., Bosse, T., Neerincx, M. A., & Cremers, A. H. M. (2017). A Digital Coach That Provides Affective And Social Learning Support to Low-Literate Learners. IEEE Transactions on Learning Technologies **11**, 1 (2017) [183].

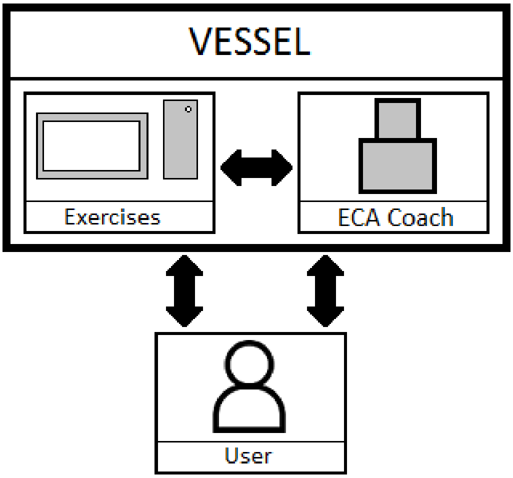


Figure 6.1: VESSEL design. System interactions are indicated with arrows: the user performs exercises, the coach monitors exercise state and user-system interaction, and the coach supports the user [217].

6.1. Introduction

In earlier studies, we have highlighted the problems that people of low-literacy encounter when trying to participate in information societies [78, 156]. Low information (reading and writing) and communication (speaking and understanding) skills cause participation issues that can be of a cognitive nature (skill application and general societal knowledge), affective nature (emotional responses like shame and fear, and low self-efficacy), and/or social nature (motivation to participate and trusting peers and teachers). We want to address these issues by designing interactive, situated societal participation learning that is grounded in crucial practical situations, which are real-life scenarios that describe the skills and knowledge needed for independent societal participation [89]. The aim is to make learning more effective, which means making the learning process more accessible (by removing or lowering barriers to entry) and making the learning experience more positive (ensuring that learners both can and want to interact with the learning), thereby supporting learners in reaching desired learning outcomes [156]. Specifically, we are designing the system *VESSEL*: a *Virtual Environment to Support the Societal participation Education of Low-literates*. VESSEL is envisioned as a set of interactive exercises grounded in the aforementioned crucial practical situations, and an autonomous, rules-driven *Embodied Conversational Agent* (*ECA*) coach that helps low-literate learners carry out these exercises by offering cognitive, affective, and social learning support. Fig. 6.1 shows a schematic VESSEL design.

We use the *Socio-Cognitive Engineering* method in the VESSEL development process (*SCE*, cf. [84, 85, 163]). This iterative software design and development method consists of three stages. In the *foundation* stage, relevant *operational demands* (actors, activities, and context-of-use), *human factors data* (theory relevant to user-system interaction), and *technology* are collected. In the *specification* stage a *requirements baseline* is created, consisting of *functional requirements* (the system's intended functionality), *claims* (hypotheses that describe the system's intended effects), *system objectives* (the system's operational or domain goals), and *use cases* (action sequences that describe the system's ideal working procedure). In the *evaluation* stage, this requirements baseline is experimentally validated.

In prior studies we have designed, developed, and evaluated two VESSEL prototypes. The first prototype [217] was a proof-of-concept consisting of four information-and-communication-skill exercises (easy and hard variants of 'online banking' and 'service desk conversation' exercises) and an ECA coach offering three kinds of learning support: cognitive support based on scaffolding (a learning support method that provides the right amount of help at the right time, [204]), affective support based on motivational interviewing (a counseling technique focused on enacting behavioural change, [207]), and social support based on small talk (a form of social interaction that is important to building interpersonal trust, [153]). All support was given as pre-recorded spoken utterances, controlled by a Wizard-of-Oz operator (cf. [179]). We evaluated this prototype to test the general applicability of cognitive, affective, and social support offered by an ECA coach. Results showed that the ECA coach improved the learning experience in all facets (cognitively, affectively, and socially), and raised learner self-efficacy regarding challenging online banking situations. Based on these positive results, a design specification for coach-driven cognitive learning support was drafted, and translated into a second prototype [246]. This work consisted of three challenging online banking exercises, and an ECA coach offering cognitive learning support based on formal scaffolding theory (cf. [227]) while following strict speech recognition rules. We evaluated this prototype (still controlled via the Wizard-of-Oz method) to test our claims of benefit with regard to cognitive support. Results showed that learner self-efficacy regarding challenging online banking was again raised, and that the formalized coach did not negatively impact the learning experience: expectations were that participants would try to interact with the coach as if it was human (i.e. asking complex questions and expecting the coach to have an answer for every situation), and that the coach's limited knowledge and strict speech recognition could cause difficulty and frustration. But this did not happen.

Now, we want to extend the VESSEL specification by also incorporating affective and social support into the design specification, thus bringing system functionality in line with the envisioned functionality from the proof-of-concept. However, trying to do so illustrates two important knowledge questions about VESSEL's ECA coach. First, we need to know how to design affective and social learning support for low-literate learners, in particular how to translate motivational interviewing theory (for affective support) and small talk theory (for social support) into support rules for the coach. Second, affective support specifically depends on understanding the learner's emotional state. We need to know what technology would allow the ECA coach to perceive and react to learner emotions, and which emotional models we can use to categorize these. As in [246], we can answer these questions by incorporating new theory into the SCE foundation of VESSEL. We update operational demands by designing one or more new scenario-based exercises that demand cognitive, affective, and social support. We update human factors knowledge by incorporating theory on motivational interviewing, small talk, and emotional models. We update technology by describing both current technology for autonomous emotion detection, and the envisioned role of this technology in VESSEL.

In summary, in this work we aim to design and evaluate a third VESSEL prototype that offers cognitive, affective, and social learning support. Four steps are needed. First, we expand the VESSEL foundation as described. Second, we refine the VESSEL design specification by operationalizing the foundation theory into comprehensive coach behavioural rules, updating the requirements baseline, and writing new uses cases. Third, we design and develop our third VESSEL prototype on the basis of this specification. Finally, we experimentally evaluate the prototype with low-literate learners. Specifically, we investigate how the new prototype affects the cognitive, affective, and social learning experience and learning outcomes of a volunteer work learning exercise, compared to our previous prototype (see [246]). This lets us answer the following research questions:

- **Q1. Design.** How can we create a design specification for VESSEL that incorporates rules for cognitive, affective, and social learning support provided by an ECA coach?
  - **Q1a.** Which emotional models, motivational interviewing rules, small talk scenarios, and measurement methods are needed to create these rules?
  - **Q1b.** Which functionalities, interaction methods, and appearances should the ECA coach have to reflect this?
- **Q2. Evaluation.** Does an ECA coach created in accordance with this specification result in a higher learning effectiveness for low-literate learners than an ECA coach that incorporates only formalized cognitive learning support?

The structure of this paper is as follows. In section 6.2, the SCE foundation is updated to address the knowledge gaps: what coach behaviour rules can be derived from motivational interviewing and small talk theory, which formal models of emotion can be used by the ECA coach, and what technological options are there for autonomous emotion detection? This information is incorporated into the SCE foundation in section 6.3. In section 6.4, the new VESSEL prototype is designed and developed. Section 6.5 describes the design and setup of the experiment created to evaluate the effectiveness of the prototype, and section 6.6 presents evaluation results. Finally, section 6.7 presents conclusions and directions for future work.

6.2. Foundation

6.2.1. Operational Demands: Exercises

To provide the right context-of-use for the envisioned cognitive, affective, and social support coach, exercises are needed that pose cognitive, affective, and social challenges (in tandem). None of our previous exercises (cf. [217, 246]) meet this

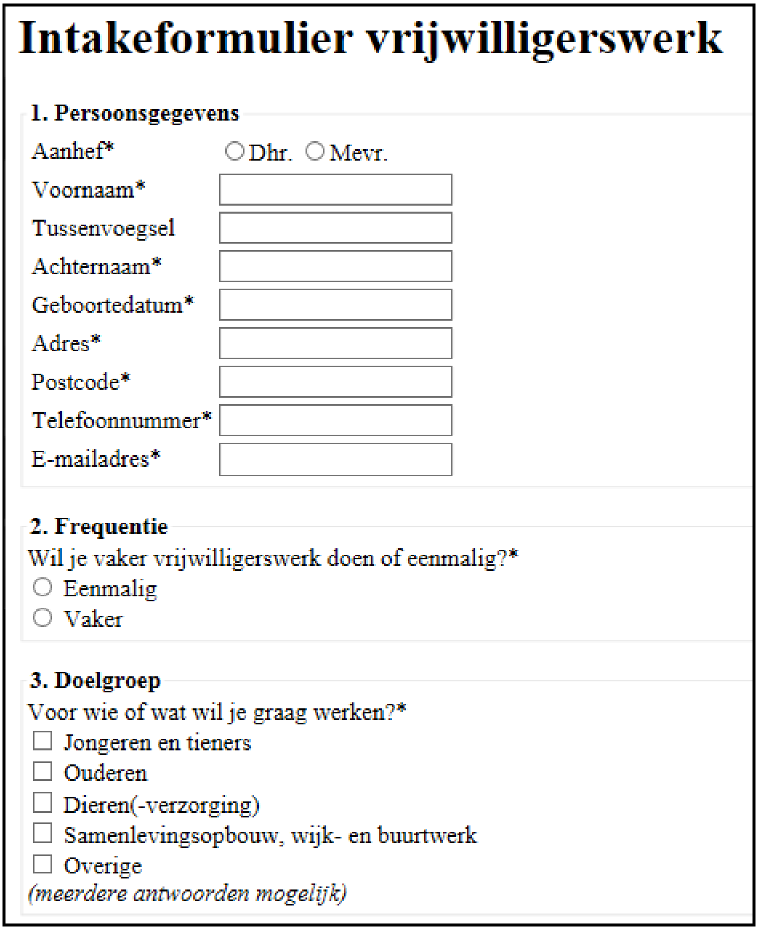
Figure 6.2: The two appearance options for the ECA recruiter.

demand. We have chosen to design a new exercise, based on the crucial practical situation 'registering for volunteer work'. The exercise consists of two parts. In the 'form' part, learners must fill out an 'intake form' for volunteer work. The form has a section for demographic information, and four sections that categorize the learner's wishes with regard to volunteer work: frequency, target demographic(s), target area(s), and useful skills possessed by the learner. This part of the exercise tests reading and language comprehension, as well as ICT skills, and presents mostly potential cognitive problems (related to vocabulary and comprehension), but also affective ones (willingness to admit interests, uncertainty about what this information is used for). In the 'recruiter' part, learners must speak to an ECA playing the role of a volunteer work recruiter. The recruiter asks a number of questions, drawn from a large set, that reference their choices on the form. Learners talk to the recruiter directly. This part of the exercise tests speaking skills and comprehension of spoken language, and presents mostly potential affective problems (fear and shame about discussing personal desires and limitations) and social ones (speaking to a formal-looking stranger about unfamiliar topics). Combined, we think that cognitive, affective, and social challenges will be presented throughout the exercise, providing room for the coach ECA to support learners in all three areas. Two versions of the exercise have been made: the order of information elements and some of the contents are different in the forms, and the recruiter ECAs are visually slightly different. Fig. 6.2 shows the two appearances of the recruiter ECA. Fig. 6.3 shows an excerpt of one exercise form.

6.2.2. Human Factors Knowledge

Emotion Models

To design an ECA coach that can give accurate affective support, we need a way to categorize and assess the intensity of learner emotions. Three general approaches



Intakeformulier vrijwilligerswerk

1. Persoonsgegevens

Aanhef* ○ Dhr. ○ Mevr.
Voornaam*
Tussenvoegsel
Achternaam*
Geboortedatum*
Adres*
Postcode*
Telefoonnummer*
E-mailadres*

2. Frequentie

Wil je vaker vrijwilligerswerk doen of eenmalig?*
○ Eenmalig
○ Vaker

3. Doelgroep

Voor wie of wat wil je graag werken?*
☐ Jongeren en tieners
☐ Ouderen
☐ Dieren(-verzorging)
☐ Samenlevingsopbouw, wijk- en buurtwerk
☐ Overige
(meerdere antwoorden mogelijk)

Figure 6.3: Excerpt of one variant of the intake form, containing questions about: demographic information, frequency of volunteer work, and intended target group to do volunteer work for.

to emotional modeling exist: the basic emotions approach, the cognitive appraisal approach, and the dimensional approach [247]. The *basic emotions* approach claims that certain core emotions are biologically based and genetically coded [248], and that emotions have evolved to increase odds of survival. Ekman [249] posits that the emotions anger, disgust, fear, enjoyment, sadness, and surprise have universal facial expressions associated with them; this makes these emotions basic emotions. Emotions without universal facial expressions (such as awe, excitement, shame, and relief) are conceptualized as a blend of these six. Similar models by Plutchik [250] and Parrott [251] categorize all emotions as primary, secondary, or tertiary emotions. The basic emotions approach is useful for VESSEL because it provides a discrete, easy-to-interpret classification of emotions. However, this classification does not allow for differentiation in the intensity of emotions. This might make this approach too broad-strokes to allow for individually tailored emotional support.

The *cognitive appraisal* approach describes an emotional state as a reaction to an arousing situation [252]. The experience of 'emotion' is attributed to physiological changes in the body [253]. By categorizing which physiological changes respond to which emotional reactions, it becomes possible to measure emotions objectively: for instance, the Facial Action Coding System (FACS, [254] is a systematic analysis of the emotions associated with facial expressions. This possibility makes the cognitive appraisal approach potentially useful for VESSEL. However, appraisals of events and the associated emotional reactions are individually and culturally variable: different people interpret body signals differently, and different cultures consider some emotions as undesirable or unacceptable [255]. Consequently, using this approach necessitates a careful study of the intended user demographic.

6

Finally, the *dimensional* approach posits that emotions are not independent discrete states, but rather that all emotions are related in a systematic manner [256]. For instance, Russel's [257] circumplex model of affect classifies emotions on the axes of valence (how positive or negative the emotion is felt as) and arousal (how excited or calm the emotion is felt as). This approach is useful for VESSEL because it allows for emotional responses with different intensities: for instance, a person can be a little bit happy because the weather outside is nice, very happy but relatively calm when spending time with family and friends, or incredibly happy and excited for winning the lottery. However, no single 'best' classification model exists. Both Oliveira, Teixeira, Fonseca & Oliveira [258] and Lewis, Critchley, Rotshtein & Dolan [259] claim that the valence and arousal dimensions are actually correlated, and cannot be treated as independent. Others have posited that a third dimension, dominance (how dominant or submissive the emotion is felt as) is necessary to adequately describe the emotion space [244, 260].

In our VESSEL design, we apply a combination of the basic and dimensional models. We define that the coach can categorize four basic emotions: anger, fear, sadness, and happiness. Based on [156], we think these emotions will play a role in our volunteer work scenario. Low-literate people experience sadness or anger when confronted with challenging information tasks, like the complex wording of the form or the difficult vocabulary of the conversation partner. They experience

Emotion	Valence	Arousal	Dominance
Anger	Negative	High	High
Fear	Negative	High	Low
Sadness	Negative	Low	*Either*
Happiness	Positive	*Either*	*Either*

Table 6.1: Categorization of four basic ECA coach emotions.

fear when confronted with decisions they feel they cannot oversee the scope of, like being asked to commit to volunteer work. They experience happiness when completing challenging tasks, particularly related to literacy. Additionally, these four emotions can easily be categorized using the three dimensional terms used by Bradley & Lang [244]: Valence/pleasure, arousal, and dominance (see Table 6.1). These two models describe all the affective functionality we want in our coach: a simple categorization of emotions that a digital coach can recognize, and a division in measurable quantities that can be used for decision purposes. In theory, the cognitive appraisal model could be used to fine-tune the Table 6.1 classification of basic emotions to low-literate people, and result in a more accurate description of how these emotions are expressed (how strongly, and in what ways). We leave this time- and labor-intensive adaptation out of our current model, and defer it to future work.

Motivational Interviewing

Motivational interviewing is originally a counseling technique aimed at enhancing an individual's intrinsic motivation to change behaviour [207, 261, 262]. The technique has also been used to provide learning support, by making learners feel good about the process, and reframing and reinforcing positive self-efficacy information [209, 211]. The motivational interviewing process consists of three strategies: affirmation, awareness, and alternatives [261]. *Affirmation* aims to establish empathy between counselor and client. This is often combined with *reflective listening* [207, 263] to put the focus of the conversation on the client's perspective, not the counselor's, motivating the client to explore their own thoughts. *Awareness* aims to help clients become aware of their problem through their own reasoning process. The *alternatives* strategy focuses on helping clients evaluate alternatives to their current situation. Sobell & Sobell [211] add two more strategies: *normalizing* aims to communicate to clients that many other people share their problems and their difficulties to change, and *self-efficacy supporting* focuses on raising the client's self-efficacy about being able to make the change.

In VESSEL, we use four of these strategies to formalize the coach's affective support. The awareness and alternatives strategies are most designed for behavioural change therapy, and are therefore less useful in a learning support setting. The remaining strategies (reflective listening, normalizing, affirmation, and self-efficacy supporting) are used to create a four-tiered model of motivational interviewing utterances. By using the four strategies in the orders presented, the coach provides affective support in a standardized way. We further specify that the coach can identify learner emotional states at three levels of accuracy: *General*, *Specific*, and *Very Specific*. If the coach identifies that the learner is in some negative-valence emotional state, but cannot the exact state, the General level of support is used. If the coach can identify the exact emotional state, the Specific level is used. If the coach can even estimate what the exact trigger is for this emotional state (a particular difficult exercise element or challenge), the Very Specific level is used. Table 6.2 provides an overview of this model, with example utterances for each category and level.

Small Talk

According to Bickmore & Cassell [40], an essential aspect of human-system interaction is building trust between user and application. They show that, in interactive systems, "*...embodied conversational agents are ideally suited for this task* [i.e. building trust] *given the myriad cues available to them for signaling trustworthiness*" [40, p. 396]. In learning, trust makes learners more receptive to teacher suggestions, and motivates learner persistence [40, 212]. Small talk is often used to establish trust. Cassell & Bickmore [153] show that small talk leads to trust-building in three ways. First, small talk establishes *solidarity*, demonstrates reciprocal appreciation, and avoids 'face threat', both because the speakers show interest in one another and because the conversation is kept on a safe level of depth. Second, it establishes *familiarity* and common ground, because speakers discuss a clearly established and accessibly topic. Third, small talk increases *coordination* between speakers, both verbally and nonverbally, as speakers must pay attention to each other and take turns talking.

In VESSEL, we use these three characteristics of small talk to write an introductory small talk session for the exercise, wherein the coach discusses the topic of volunteer work with the learner. The session consists of a number of possible phrases and questions that the coach can say, ordered in a particular way to ensure a logical conversation flow; see Appendix G for an overview of this. To establish familiarity, the coach only discusses the established topic of small talk. To evoke solidarity, the coach both asks the user about their volunteer work experiences, and talks about their own 'experiences with volunteer work'. The coach asks follow-up questions whenever possible, but does not push learners if they are not interested in answering. To establish coordination, the coach follows a simple operation schema: whenever the learner starts talking after a question, the coach does not interrupt. Whenever the learner stops speaking, the coach waits three seconds, then utters the next phrase or question that makes sense in the scenario.

6.2.3. Technology: Emotion Measurement Tools

Emotion measurement tools can be grouped in three categories: Psychological, physiological, and behavioural [264]. *Psychological* tools are subjective self-report tools, such as questionnaires. These tools are inexpensive, unobtrusive, and non-invasive, and they are the only way to measure a participant's inner perception and experience [265]. However, language barriers and cultural differences in emo-

Description	General	Specific	Very Specific
Reflective listening. This utterance makes explicit what emotional state the coach is perceiving, and (if applicable) the issue that's causing this state. This is put in the form of a statement, not a question. The learner has a chance to provide feedback if the coach's read is incorrect.	"It looks like you are experiencing difficulties."	"It looks like you are afraid."	"It looks like you are afraid of what could happen, if you fill out this form incorrectly."
Normalizing. This utterance puts the learner's issue and emotional reaction in a broader context, to show them that they are not alone.	"Many people encounter these difficulties."	"Many people become afraid in these circumstances."	"Many people become afraid in these circumstances."
Affirmation. This utterance tells the learner that the coach understands their emotional reaction, which is 'normal' (i.e. not exceptional or strange).The coach then helps the learner move look forward, by suggesting an action they can take, reminding them about help they can receive, or giving moral support.	"It is not strange that this is challenging for you. With practice, you will get better."	"It is not unusual for you to be afraid here. Keep trying, and you will see that it is not as difficult as you think."	"It is not unusual for you to be afraid of this. But this form is only a first step. In the interview afterwards, you will be able to clarify what volunteer work you do or do not want to do."
Self-efficacy supporting. This utterance tries to raise the learner's self-efficacy regarding the exercise topic and/or their skill in doing the exercise.	"I think, that you have already achieved a lot today." today."	"I think, that you have already achieved a lot	"I think, that you have already achieved a lot today."

Table 6.2: Categorization of four basic ECA coach emotions.

6

tion (cf. [255]) can make results unreliable. Participants may also be unwilling to talk about their emotional state to researchers, particularly in embarrassing cases, or they may be unable to put their emotional state to words. Finally, emotional self-reporting can be difficult in parallel with an experimental task without causing interference [265]. In practice, these tools can only be used after experimental sessions or in-between exercises.

Physiological tools are objective measures that use sensors. For instance, heart rate and galvanic skin response can be measured to determine arousal. Sensors can provide a continuous objective monitoring of the person's state [264] without being disruptive of task performance [266]. However, sensors can be invasive or intrusive, which could potentially influence the user's experience [265]. Sensors also often require specialized equipment and technical expertise to be used correctly. Using them sub-optimally, or in the presence of confounding circumstances such as excessive lighting or heat, may result in noisy data [264].

Lastly, *behavioural* tools measure motor-behavioural expressions and changes in physiological state. Unlike physiological tools, which are directly interested in the state of the body, behaviour tools measure body state in order to assess behaviour. Commonly, non-intrusive devices like computers and microphones are used: Zimmermann, Guttormsen, Danuser & Gomez [267] describe an example where "[the method] *extracts motor-behavioral parameters from log-files of mouse and keyboard actions, which can be used to analyze correlations with affective state*" [267, p. 540]. The user's actions and behaviour can be used to predict and assign valence and arousal scores [267]. This approach is not very invasive (but [265] notes that participants consider video cameras to be obtrusive), doesn't interfere with task performance, and can detect emotional cues that other tools cannot measure, such as facial expressions. However, special hardware and software are needed to capture this kind of data [268], and interpreting it requires trained, experienced and objective observers [264]. Additionally, the interpretation methods are commonly tested on 'produced' emotional expressions; in natural situations, recognition accuracy can drop harshly in case of spontaneous emotions [265].

In VESSEL, we combine all three tool types, to make use of the advantages of each. Questionnaires are administered at the start and end of each exercise to gauge users' self-reported affective state. We expect all participants to fill out these questionnaires, making this a reliable source of data that will be useful for statistical evaluation of VESSEL's affective learning experience. Physiological and behavioural tools are used during exercises. The Shimmer sensor package [269] is used to measure learner arousal, using a photoplethysmographic (PPG) sensor attached to the earlobe (see 6.5). PPG sensors measure changes in light absorption that result from subcutaneous blood flow, which is translated into a measure of heart rate. The FaceReader facial recognition software package [270] uses a webcam to capture video of the learner's face, measure learner valence, and attempt to identify the occurrence of six basic emotions: happiness, anger, sadness, fear, surprise, and disgust (the latter two of which we are not interested in). Both body sensors and facial recognition let us rapidly assess users' affective states, and offer immediate affective support that is accurately tailored to that state; the FaceReader provides more detailed evaluation of the user's affective state, making it especially useful for support provision, while the Shimmer's arousal detection is a more objective measure of participant physiological state over time that may also be useful for later quantitative analysis.

6.3. Specification

6.3.1. Operationalization

We take three steps to operationalize the cognitive, affective, and social support behaviour for our prototype ECA coach. As a baseline for the prototype, we adopt the cognitive support model created in Massink [246]. This model contains a systematic way to create cognitive support utterances to cover all potential cognitive difficulties in an exercise, a comprehensive set of behaviour rules for an ECA coach to provide cognitive support, and rules to model speech recognition. Cognitive support utterances are created for all identified 'difficult elements' in the exercise, for each of Massink's [246] five levels of cognitive support. We incorporate the rules for when and how to provide cognitive support into this prototype as they are. Since cognitive support is not evaluated with this prototype, the full process is left out of this paper (but interested readers are referred to [246]). Similarly, we apply the techniques for emulating speech recognition: we create a dictionary of keywords (based on exercise 'difficult elements' and cognitive support categories), and define how the coach can react to these keywords. We make one significant change: during the recruiter part of the exercise, cognitive support utterances will be spoken by the recruiter ECA, not the coach ECA. This change is made because our earlier work with two concurrent ECAs [217] shows that asking a coach character for help in a conversation exercise interrupts the dialogue, and leads to learner confusion.

To operationalize affective support, we use the four motivational interviewing categories and three specificity levels from Table 6.2 to create affective support utterances: for every emotion the coach ECA can recognize except happiness (which does not require support), we create one or two utterances for every category-level combination. We then define when and how these utterances are used. We say that affective support must be given in three circumstances. If learner arousal is high but valence is not clearly low, or if valence is low but arousal is not clearly high, the coach detects that affective issues are happening, but cannot quantify which. In this case, General support is given (as per Table 6.2). If arousal and valence clearly indicate anger, fear, or sadness as per Table 6.1, or if the FaceReader program strongly detects anger, fear, or sadness, the coach gives affective support on the Specific level, tailored to that emotion. If the cause of the anger, fear, or sadness is also clearly detected, the coach gives affective support on the Very Specific level, tailored to that emotion and cause. We have identified a number of elements and situations in the exercise that will likely lead to particular affective reactions, for which Very Specific affective support utterances were recorded. For example, when the recruiter ECA very curtly asks questions, we suspect low-literate participants will get angry at this disrespectful style of speaking. When giving affective support, the coach uses four utterances in sequence: reflective listening, normalization, affirmation, and self-efficacy supporting. All utterances are 5 seconds apart. After giving all four types of affective support utterance in a row, the coach must wait at least one minute before giving more: this is done to prevent endless repetition of the same support for learners that stay in the same affective state for a longer time.

Finally, we operationalize social support by using the small talk utterance corpus in Appendix G. The coach uses the utterances as indicated. We define the speech recognition options for small talk here: when the coach asks the learner a question, it can understand all varieties of 'yes' or 'no' as answers, and react accordingly. As long as a learner is talking (to answer a question, or for other reasons), the coach recognizes this and does not talk or interrupt. When the learner is not talking, the coach moves through small talk utterances, keeping 5 seconds between each.

6.3.2. Requirements Baseline

Section 6.2's foundation data are now used to refine the existing VESSEL requirements baseline. Only those requirements that change on the basis of the expanded foundation are refined, for the *coach* and *exercises* aspects of VESSEL; requirements that are not described do not change. Table F.1 (see Appendix F) shows the refined requirements baseline.

Requirement **R1. Adaptability** is refined for both coach and exercises. The coach (**R1.1-C**) should ensure that affective support matches the learner's emotional state. Affective support must only be given if the sensors indicate particular emotional valence and intensity. The exercises (**R1.1-E**) should be cognitively and affectively challenging. An exercise is affectively challenging if learners experience significant anger, fear, or sadness at least once while doing the exercise.

Requirement **R2. Sensitivity** is refined for the exercises (**R2.1-E**), as an extension of **R1.1-E**. Exercises must be as sensitive *or insensitive* as needed to reach intended difficulty levels. Specifically, in exercises that feature conversation partners, the conversation partner's dialogue must display the right level of sensitivity to effect the intended affective difficulty. If the conversation partner is too kind, no affective difficulty is reached (see [217] for an example of this), but if the conversation partner is too abrasive, low-literate users might stop the exercise midway [156].

Requirement **R6. Support** is zoomed in to coach-offered affective support (**R6.2-C**) and social support (**R6.3-C**). The coach should offer affective support and social support according to the behaviour rules in section 6.3.1. Social support should be offered before the exercises, and affective support should be offered during the exercises, concurrent with cognitive support (**R6.1-C**, remaining unchanged). No support is offered after exercises.

Requirement **R7. Interactivity** is refined for the coach and the exercises. The coach's proactive affective support interactions with the learners (**R7.1-C**) should be driven by sensors and facial recognition, and all proactive support interactions should be guided by the rules in section 6.3.1. If (during an exercise) cognitive learning support is offered in an exercise that has a conversation partner present (*such as the recruiter ECA*), the utterance should be spoken by the conversation

6

Figure 6.4: The two appearance options for the ECA coach.

partner ECA instead of the coach ECA (**R7.1-E**). This applies to both proactive and reactive cognitive support.

6.4. Evaluation: Prototype Development

Functionality. The prototype consists of the two-part volunteer work exercise described in section 6.2.1, and an ECA coach that offers cognitive, affective, and social learning support according to the rules and timing approaching described in section 6.3.1.

Interaction methods. Learners interact with the form part of the exercise using mouse and keyboard. They can talk to the coach and the recruiter in natural language. For the purposes of evaluation, coach and recruiter are designed to be controlled via the Wizard-of-Oz method [179] similar to Massink [246], meaning that speech recognition is emulated by the Wizard operators. This must be done in accordance with the speech recognition rules in section 6.3.1.

Appearance. The coach ECA avatars were developed in Unity. Two visual variations of the coach were made to match the two variant exercises (section 6.2.1). Both were based on the appearance of our previous coaches [217, 246]. Fig. 6.4 shows the two coach appearances. The recruiter ECA avatars (Fig. 6.2) were also created in Unity.

6.5. Evaluation: Methods

6.5.1. Experimental Design

An experiment was designed to evaluate the learning effectiveness of our VESSEL prototype. We wanted to compare this prototype, which offers 'full' (cognitive, affective, and social) learning support, to a prototype that offers only cognitive learning support, built according to our previous specification [246]. Six hypotheses were created, corresponding to the six learning effectiveness outlined earlier: cognitive, affective, and social learning experience, and cognitive, affective, and social learning outcomes. In general, we expect that the current 'Full Support' prototype results in higher learning effectiveness on all fronts than the 'Cognitive Support' prototype.

- **H1. Cognitive Experience (Performance).** Learners that receive full learning support report better performance during the exercise, expend less effort doing the exercise, complete the exercise quicker, and receive more support while doing the exercise, than learners that receive only cognitive learning support.
- **H2. Affective Experience (Positive Affect).** Learners that receive full learning support report a more positive affective state than learners that receive only cognitive learning support.
- **H3. Social Experience (Motivation).** Learners that receive full learning support are more motivated to learn and to continue learning than learners that receive only cognitive learning support.
- **H4. Cognitive Outcomes (Success).** Learners that receive full learning support remember and recall more details about the exercise than learners that receive only cognitive learning support.
- **H5. Affective Outcomes (Self-Efficacy).** Learners that receive full learning support report a higher increase in self-efficacy than learners that receive only cognitive learning support.
- **H6. Social Outcomes (Coach Opinion).** Learners that receive full learning support have a more positive view about the coach, and initiate more interactions with the coach, than learners that receive only cognitive learning support.

To evaluate these hypotheses, we designed a mixed-method repeated-measures experiment that combined within-subjects and between-subjects measurements. The main independent variable was **Support Model**, with two levels: *Full Model*, and *Cognitive Model*. Participants were invited to work with both prototypes: in one experimental session, participants completed the exercise twice, once with the full prototype and once with the cognitive prototype. Prototype order was counterbalanced: 50% of participants did the Full Model condition first and the Cognitive Model condition second, and 50% did the opposite. The two versions of the coach and the two exercises were counterbalanced as well, leading to eight different orders.

6.5.2. Measures

Seventeen quantitative dependent variables were measured. Twelve were self-report questions, measured using three questionnaires (section 6.5.4), and five were objective performance metrics. Appendix I shows the variables.

6.5.3. Participants

Participants were recruited from five reading and writing classes throughout the Netherlands (located in Rotterdam, Nijmegen, and Den Helder). We used Kurvers et al.'s [89] five language learner profiles to select participants for this study. Only learners that matched profiles 2, 3, and 4 were invited: learners in profiles 1 and 5 are respectively too skilled to benefit from our level of exercise and support, and too low-skilled to independently engage with the language level and complexity level of our prototype. Thirty-four participants completed the entire experiment: twenty men and fourteen women, with ages ranging from 19 to 64 (*M=41.3, SD=15.1*). Ten participants self-identified as being natively fluent in Dutch. The other twenty-four identified as 'somewhat fluent'. Other languages spoken, either natively or as a second language, included: Arabic, Amharic, Aramaic, Bosnian, Catalan, Dari, Edo, English, Farsi, French, Italian, Moroccan, Papiamentu, Russian, Somali, Spanish, Swahili, Swedish, Tamil, and Turkish. Twenty participants reported having prior experience with volunteer work.

6.5.4. Materials

The experimental setup consisted of two laptops, each connected to one monitor (Fig. 6.6), which were used to run the experiment. The monitors allowed participants to see and interact with the exercises. The laptops allowed experimenters to run the exercise (left laptop) and the coach (right laptop). On the participant side, a mouse, keyboard, speakers, and microphone were provided: mouse and keyboard let participants fill out the form, speakers played the coach and recruiter ECA utterances, and the microphone was used to suggest that participants could talk to the ECA characters, as well as to record audio (with consent). A webcam was attached to the left monitor, to capture visual data for the FaceReader software: participants were told this allowed the coach to 'see' them. One Shimmer sensor was used (Fig. 6.5).

Three questionnaires were used. Two questionnaires measured the fifteen self-report variables shown in Appendix I. The 'participant assessment questionnaire' (PAQ) measured participant self-efficacy on four topics (reading Dutch, computer use, filling out a form, and having a conversation in Dutch), participant motivation to learn, participant fear of going to school, and participant affective state on the dimensions of valence, arousal, and dominance. The 'exercise reflection questionnaire' (ERQ) measured participants' view on their performance and exercise results, as well as their view of the coach. Two answer methods were used. Answers to the three participant affective state questions (PAQ.7, PAQ.8, and PAQ.9 in Appendix

Figure 6.5: Shimmer sensor with PPG ear clip.

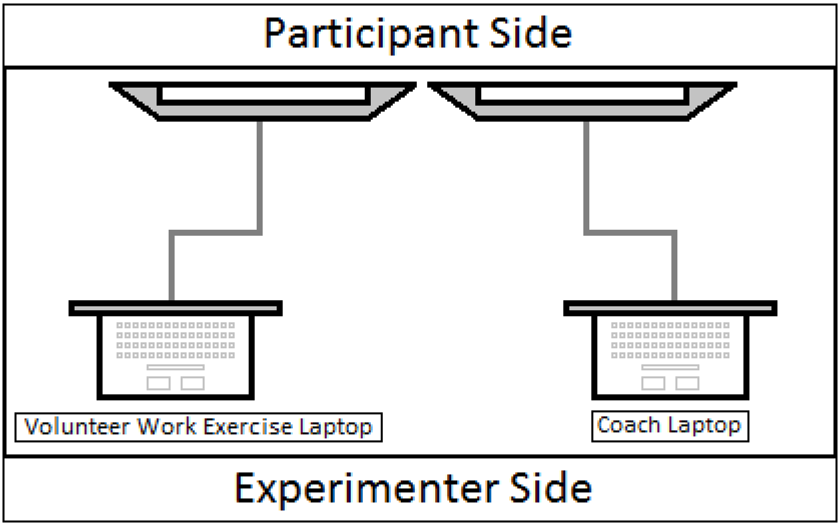


Figure 6.6: Schematic overview of experimental setup. Two laptops (bottom figures) connect to two monitors (top figures). Located near the monitors are also: a keyboard, a mouse, a microphone, and a webcam (not shown in image).

6

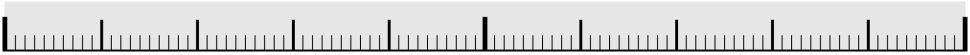


Figure 6.7: Visual analogue scale used in the PAQ and ERQ.

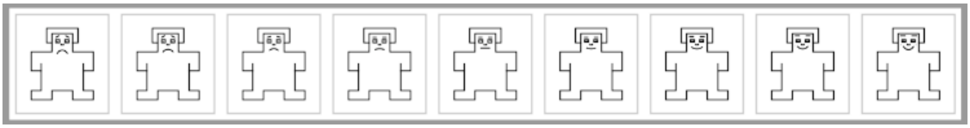


Figure 6.8: Self-Assessment Manikin bar used to measure PAQ.7.

I) were given using the Self-Assessment Manikin (Fig. 6.8, cf. [244]). Answers to all other questions were given using a visual analogue scale (Fig. 6.7). Questions were read aloud to participants, who would then mark answers on the corresponding bar; this method ensures that participant reading and writing skills are not a factor in accurate answering. The fourth 'demographic' questionnaire measured: participant age, sex, time period lived in the Netherlands, known languages, and prior volunteer work experience.

For objective measures, exercise completion time was measured using a digital clock and a stopwatch. The number of coach support utterances received by participants was recorded by hand, and categorized in the following way: utterances were either cognitive or affective support utterances (social support utterances were not recorded), they were recorded during the form part or the recruiter part of the exercise, and they were initiated either by the coach or by the participant. Finally, a 'recall test' was created to measure learning success. After each exercise, participants were given one minute to name as many form elements as they could remember. Researchers wrote down which of the five categories on the form (see Fig. 6.3) participants named. Score was calculated per category: 1 point if the category was named and described correctly, or 0.5 points if either the category was named correctly, but not described, or if it was described (ex. by giving examples of category contents) but not named, up to a maximum score of 5 points.

Note here that discrete participant emotional states as measured by the Shimmer and FaceReader (happiness, anger, sadness, and fear) are not used for hypothesis evaluation. As described in section 6.2.2, this basic model of emotions is useful for driving immediate coach decisions. We instead use the Self-Assessment Manikin (Fig. 6.8), which is an expression of the dimensional approach, as we believe this allows for more in-depth analysis of emotional states.

6.5.5. Procedure

Each thirty-minute session started with a general introduction, informed consent forms, and the demographic questionnaire. Researchers explained the general experiment flow and the experimental setup hardware (Fig. 6.6). The Shimmer sensor was introduced and attached to the participant's earlobe: doing this before measurements were taken gave researchers time to calibrate the exact placement for optimal results, and allowed participants to get used to the sensation. The first

PAQ was administered. Researchers then activated the designated coach prototype. The ECA coach (controlled Wizard-of-Oz style by one researcher who followed the control rules described in section 6.3.1) introduced itself to the user and explained the first exercise. Participants were told to complete the exercise in two steps: fill out the form, then have a conversation with the recruiter. Participants were given as much time as needed to complete the exercise. After the first exercise, the first ERQ and the second PAQ were administered, followed by a first recall test. Researchers then activated the other coach prototype, after which the second exercise was introduced and conducted similar to the first. After the second exercise, a second ERQ and third PAQ were administered, and a second recall test. Finally, participants were fully debriefed (including a behind-the-scenes look of the VESSEL prototype and the Wizard-of-Of method) and rewarded for participation.

6.6. Evaluation: Results

Four analysis steps are presented here. Section 6.6.1 shows the evaluation of the six main hypotheses (section 6.5.1). Section 6.6.2 looks at potential order and learning effects between the two exercises. Section 6.6.3 describes qualitative observations made by the researchers, during the experiment and by listening to recorded audio proceedings afterwards. Based on observations and initial results, section 6.6.4 shows two post-hoc analyses. Before analysis, the data was characterized and checked for irregularities. No obvious mistakes or irregularities were found. Questionnaire reliabilities were assessed: Cronbach's alpha was .773 for the PAQ (.810 if based on standardized items) and .872 for the ERQ (.844 for standardized items). No data reduction measures were used.

6.6.1. Hypothesis Evaluation

To evaluate the six hypotheses data from the PAQ, ERQ, and DM data were subjected to repeated measures General Linear Model (GLM) analyses. The ERQ and DM data were analyzed with one two-level factor: **Coach Type**, with levels 'Full Coach' and 'Cognitive Coach', The PAQ data were analyzed with one three-level factor: **Coach Type**, with levels 'Before Exercise', 'Full Coach' and 'Cognitive Coach'. Three significant results were found. In the Full Coach condition, participants received more affective support than participants in the Cognitive Coach condition, in both form (F=14.431, p=.001, β=.957) and recruiter (*F=52.755, p=.000, β=1.000*) parts, as well as more 'total' support (cognitive and affective support combined) during the form part (*F=29.005, p=.000, β=.999*). This supports H1: learners receive more support during the form part of the exercise. Also, participants in the Full Coach condition actively initiated more learner-coach interactions, i.e., asked the coach more questions without prompting, than participants in the Cognitive Coach condition (*F=8.484., p=.007, β=.806*). This supports H6: learners engage in more self-started interaction with the coach. Appendix H shows the full results.

6.6.2. Exercise Order Effects

To test for exercise order effects, the PAQ, ERQ, and DM data were used in two more repeated measures GLM analyses. Similar to the previous, the ERQ and DM data were analyzed with two-level factor (**Exercise Order**, with levels 'First Exercise' and 'Second Exercise') and the PAQ data were analyzed with one three-level factor (**Exercise Order**, with levels 'Before Exercise', 'First Exercise' and 'Second Exercise'). The following significant results were found. Self-efficacy about reading Dutch ($F=3.848$, $p=.032$, $\beta=.562$) and about volunteer work ($F=5.635$, $p=.008$, $\beta=.825$) were higher after the second exercise than after the first exercise. Participants also reported lower arousal after the second exercise ($F=4.754$, $p=.036$, $\beta=.562$). Related to the form part of the exercise, completion time ($F=16.042$, $p=.000$, $\beta=.972$), number of cognitive support utterances received ($F=8.403$, $p=.007$, $\beta=.802$), and total amount of support utterances received ($F=5.049$, $p=.032$, $\beta=.586$) were all lower after the second exercise. Finally, related to the recruiter part of the exercise, participants started more learner-coach interactions during the second exercise ($F=9.782$, $p=.004$, $\beta=.858$). Appendix H shows the full results.

6.6.3. Observations

6

Experimental observations showed that in general, learners managed to use the prototype and work with the coach as intended. Learners engaged with and completed the exercises, and they listened to and asked for help from the coach. This was particularly true for the Cognitive Support prototype, which was observed to work almost exactly like the prototype in our previous experiment (see [246]). The provided cognitive support was sufficient to help learners in both the form and recruiter parts of the exercise. As in Massink [246], learners almost always listened to coach advice, and would only occasionally ask questions themselves. In the recruiter part of the exercise, participants adapted to talking/listening to the recruiter ECA with no problems, and had no problems with the recruiter providing cognitive support. One unexpected side effect was that participants would almost 'forget about the coach', since in this prototype it had no other support to give. But this did not seem to negatively influence the relation between learner and coach, with learners mostly expressing amusement, "*Oh, she's still here too!*", whenever they noticed the coach (still visible on the right screen).

In the Full Support prototype conditions, the 'small talk' social support worked almost exactly like the small talk in our first proof-of-concept prototype [217]. Participants seemed to honestly and genuinely speak with the coach about their volunteer work experiences and preferences. Cognitive support in this condition worked similar to the Cognitive Support condition, the only difference being that researchers felt that participants actively asked more questions. However, the affective support seemed to work only piecemeal. Providing affective support was hampered because our sensors did not work as well as hoped. The Shimmer data was noisy, and prone to halting and resuming at random moments. The FaceReader data generally gave clearer reads on participants' emotional states. However, we encountered the issue that some participants had resting facial expressions that the FaceReader interpreted as a particular emotion: for instance, one participant's facial features were interpreted as a high level of 'sadness' all the time, leading to the coach repeating similar affective support every minute. Another problem was that both Shimmer and FaceReader had serious difficulty working with darker skin tones: the FaceReader algorithm was less effective at reading black and tan faces, and the Shimmer's PPG (which works by sending red light through the earlobe) seems to have been calibrated on light skin, not taking the different light absorption/reflection profile of dark skin into account. As a result, in the cases of many dark-skinned participants (who made up a significant subset of the NT2 group), we simply did not have enough accurate data to provide affective support to begin with. As a solution, we decided to incorporate the personal situational interpretation of the wizard operators into the decision making process: if both researchers agreed that a participant was clearly exhibiting a certain emotion, affective support could be provided. In practice, this agreement was not reached very often, and as a result, the provision of affective support to these participants was limited.

In practice, the Very Specific level of support was never used, as the situations we expected to necessitate this support (and recorded these utterances for) were not seen. The reception of the General and Specific support levels was mixed. Participants did often verbally acknowledge the support (for instance, by responding to or thanking the coach). And during debriefing, participants would often mention that the Full Support coach "*cared about [the participant] more*". However, sensors never showed a direct physiological reaction to affective support (that was clear enough to discern with accuracy). This makes it unclear to what degree the affective support had the intended effects. One other unexpected observation was that some participants would countermand the coach's affective support: reflective listening statements like "*It looks like you are scared*" were sometimes met with negations such as "*No I'm not.*" We added to our control rules that, in these cases, the rest of the affective support for this occurrence should be cancelled.

Finally, unexpected differences were seen between the NT1 and NT2 participant groups. The NT1 participants were generally better at the recruiter part of the exercise, due to their native Dutch speaking and large vocabulary, but worse during the form part of the exercise, due to limited ICT and computer skills. The NT2 participants showed the inverse: good computer skills, but limited Dutch vocabulary. While these differences have been seen to some degree in our earlier work (see particularly [156] for our overview of meaningful differences), this experiment marks the first time in our evaluation of VESSEL prototypes that significantly different outcomes were found between the groups (as per section 6.6.3).

6.6.4. Post-Hoc

Based on the aforementioned observations and analysis results, we decided to investigate the effect of learner background, or 'type'. Two types of learners were identified: 9 participants were learners with a native Dutch background ('NT1'), and 25 participants were learners with a migrant background ('NT2'). The previous repeated measures GLM analyses were then repeated, using **Learner Type** as a between-subjects variable. Results suggest significant differences between the experiences of the two types. NT1 learners reported higher (self-reported) performance (*F=4.585, p=.040, β=.547*), higher valence (*F=5.918, p=.021, β=.655*), less received cognitive support in both the form (*F=4.586, p=.040, β=.545*) and conversation (*F=5.350, p=.028, β=.610*) parts, and lower completion time in the recruiter part (*F=10.387, p=.004, β=.871*). NT2 learners reported higher computer use self-efficacy (*F=4.171, p=.025, β=.692*), more self-initiated coach interaction in the recruiter part (*F=8.589, p=.004, β=.850*), and a higher desire to use the coach again in the future (*F=7.508, p=.010, β=.757*). Additionally, we found one interaction effect for Learner Type and Exercise Order: NT1 learners reported spending high effort on the first exercise and low effort on the second, while NT2 learners reported moderate effort on the first exercise and high effort for the second (*F=9.888, p=004., β=.862*).

We also tested the variables of age, sex, time spent living in the Netherlands, experience with volunteer work, and counterbalancing order for between-subjects effects. Three significant effects were found for **Learner Sex**. Women received more affective support than men, overall (*F=9.333, p=.005, β=.840*). An interaction effect between Learner Sex and Coach Type showed that during the form exercises, men received a higher number of support utterances in the Cognitive Coach condition, and women received a higher number of support utterances in the Full Coach condition (*F=6.049, p=.020, β=.663*). Finally, an interaction effect between Learner Sex and Exercise Order showed that for women, self-efficacy with regard to holding a conversation was significantly higher after Exercise 1 than either at the start of the experiment, or after Exercise 2. For men, this difference did not exist (*F=3.586, p=.040, β=.621*). Appendix H shows the full results of both Learner Type and Learner Sex analyses.

6.7. Conclusions

This study aimed to answer two research questions. Question Q1 was: "*How can we create a design specification for VESSEL that incorporates rules for cognitive, affective, and social learning support provided by an ECA coach?*" This question was answered in sections 6.2 through 6.4. Sub-question Q1a, "*Which emotional models, motivational interviewing rules, small talk scenarios, and measurement methods are needed to create these rules?*", was answered in section 6.2. An overview was created of operational demands (a description of the volunteer work exercise to be used in the prototype), human factors knowledge (the three kinds of extant emotional models, and our systematic interpretations of motivational interviewing and small talk), and technology (three kinds of autonomous emotion measurement tools). This overview was incorporated into the SCE foundation of our VESSEL design specification: we created the volunteer work exercise, made rules to describe motivational interviewing and small talk behaviour, and selected the Shimmer and FaceReader sensors for use with the prototype. Sub-question Q1b, "*Which functionalities, interaction methods, and appearances should the ECA coach have to reflect this specification?*", was answered in sections 6.3, where the design specification was updated with new control rules and functional requirements, and 6.4, where a new VESSEL prototype was created based on the updated specification.

Question Q2 was: "*Does an ECA coach created in accordance with this specifi-*

cation result in a higher learning effectiveness for low-literate learners than an ECA coach that incorporates only formalized cognitive learning support?" This question was answered in sections 6.5 and 6.6, where we experimentally evaluated the prototype by comparing it against a prototype built according to our previous design specification [246]. Six hypotheses were tested:

- **H1. Cognitive Experience (Performance).** This hypothesis is partially supported. Learners in the Full Support condition did receive significantly more learning support than learners in the Cognitive Support condition, but did not report better performance, expend less effort, or complete the exercise quicker.
- **H2. Affective Experience (Positive Affect).** This hypothesis is not supported. There was no significant difference in affective state during and after the exercises between learners in either condition.
- **H3. Social Experience (Motivation).** This hypothesis is not supported. There was no significant difference in motivation to learn and to continue learning between learners in either condition.
- **H4. Cognitive Outcomes (Success).** This hypothesis is not supported. There was no significant difference in recall test results between learners in either condition.
- **H5. Affective Outcomes (Self-Efficacy).** This hypothesis is not supported. There was no significant difference in self-efficacy increase between learners in either condition.
- **H6. Social Outcomes (Coach Opinion).** This hypothesis is partially supported. Learners in the Full Support condition initiated more interactions with the coach than learners in the Cognitive Support condition, but they did not report a more positive view about the coach.

These hypothesis results seem to indicate that few differences exist between the Full Support and Cognitive Support coaches. Of the two partially supported hypotheses, H1 does not provide much information: that learners in the Full Support condition receive more support overall is easily explained by the fact that these learners received cognitive *and* affective support during the exercise, where learners in the Cognitive Support condition *only* received cognitive support. Hypothesis H6 does show an interesting finding: learners in the Full Support condition were quicker to proactively talk to the coach. This finding matches our expectation that adding affective and social support makes it clearer to learners that the coach can be talked to like a human conversation partner. In our previous work, learners that used the 'proof-of-concept' prototype (which included early operationalization of affective and social support, see [217]) were observed to proactively speak with the coach more than learners that used our first cognitive support prototype [246].

The results in this study now statistically validate these observations. It is currently unsure what mechanisms lead to this increased 'affordance of being spoken to'. Future studies could try to disentangle the effects of affective and social support, to see if either can be pinpointed as the cause, or if the effect only happens with a combination of support types.

Based on the results presented above, we must conclude that very little differences existed between the Full Support and Cognitive Support prototypes. The addition of affective and social learning support did not have many of the predicted effects. Three potential explanations are offered here. The first explanation is that our affective support manipulations may not have been large enough to produce an effect. Section 6.6.4 describes how sensor problems led to issues with the provision of affective support. In practice, the researchers only confidently employed affective support in a limited number of situations. This should be considered an oversight on our part: We expect that better results can be obtained by extensively testing and calibrating the Shimmer and FaceReader sensors to our participants. The cognitive appraisal method (described in section 6.2.2) to do could possibly be used for this. Since emotion sensing technology is still showing shortcomings in *in situ* applications like this one, future work should investigate whether or not incorporating this method in the design of affective support is valuable. Also, the volunteer work scenario used in the prototype did not seem to result in a great deal of affective challenge. Selecting an exercise for this goal is difficult: While some crucial practical situations have obvious affective impact (such as health-related issues like hospitalization, or death of a family members) it was considered ethically unjustifiable to use this level of affective stress to evaluate a digital coach. The volunteer work scenario was seen as having the potential to be affectively challenging, which we could bring out by carefully designing the difficulty and the affective and social behaviour of the recruiter ECA (see section 6.2.1). It is unclear whether or not this worked. The second explanation is that our approach to providing affective learning support with an ECA has been too limited. Studies indicate that ECAs in general can potentially change the affective experience of doing computer exercises [199] and emotionally connect to learners [201]. The *Multimodal Affective and Reactive Character* framework [271] describes three factors that influence the effectiveness of affective characters: The capacity of the agent to respond to the user in real-time, subtle visual indicators of agent affective state, and the ability to express differences in affective reactions to different individual learners. VESSEL can act in real-time and adapt to individual learners, but does not use visual indicators of affective states: The appearances and facial expressions of our coach and recruiter ECAs were more or less static (see Figs. 6.2 and 6.4). The way embodied characters look impacts how their functionality and possibilities are perceived (cf. [201]). It is possible that the more 'stylized' (non-realistic and exaggerated) appearance of our coach (Fig. 6.4) impacted this, and that a more 'naturalistic' (human-looking) appearance, including affectively expressive facial expressions, would have served us better [271, 272]. The third explanation is that our experimental setup impacted coach effectiveness. Participants used both coach versions in a span of 30 minutes: This may have caused them to see both coaches as a single entity, with slightly different behaviour between exercises. Support for this is offered by a debriefing observation: when participants were asked if they noticed any differences between the two coach ECAs, many would say yes, and then describe differences between the coach ECAs and the recruiter ECAs. In this scenario it is possible that the coach did have affective effects, but that the attribution of these effects (particularly in questionnaires) was confounded by the presence of the recruiter. Future studies should try to disentangle these effects more clearly: Maybe exercises with conversation components should use the coach as the conversation partner directly, or maybe the coach ECA should be hidden from view in these cases. In general, the lack of significant results for our affective learning support is counter to expectations (cf. [199, 201, 271, 273]), and future work in this direction should focus on investigating ways to resolve this.

Results for the order effects evaluation (section 6.6.2) show that the prototype in general did work as expected. Participants always completed all exercises using cognitive support, similar to our previous cognitive support prototype [246]. Participants accepted and understood the coach, and used its help to get through the exercises when needed. The lower completion time and lower need for learning support in the second exercise compared to the first exercise indicates a straightforward (and expected) learning effect. The lower arousal and the increased number of self-initiated learner-coach interactions in the second exercise seems to suggest that learners were more 'at ease' with the system the second time around. Finally, the increase in self-efficacy (with regard to 'reading Dutch' and 'volunteer work') over the exercises is interesting, as this reproduces our findings in Deneka [217] and Massink [246]. Appendix H shows clearly that the self-efficacy increase happened after the first exercise. Learners judge their self-efficacy lower before doing any exercise, judge it higher after completing an exercise for the first time, and then stay on that higher level throughout. One of the strongest sources of self-efficacy information is successfully completing a task yourself ('enactive mastery', cf. [109]). Results from all our prototype experiments suggest that working with VESSEL provides this: self-efficacy about a larger domain ('reading Dutch' or 'doing online banking') increases after completing a specific scenario exercise, and stays high. It would be interesting for future studies to investigate how long these self-efficacy increases last. For instance, does self-efficacy remain high after four or five exercises? And does self-efficacy remain high if longer amounts of time (i.e. weeks or months) pass between exercises?

Finally, the differences we found between NT1 and NT2 learners highlight the importance and added value of personalization. One possible explanation is that the volunteer work scenario in this prototype has caused this. This exercise is more grounded in Dutch society than online banking exercises [217, 246], and it is not a topic that both NT1 and NT2 learners have a lot of direct experience with (unlike Deneka's [217] city hall passport exercise). In this prototype's exercise, NT1 and NT2 learners encountered different problems, and showed different reactions. Combined with the aforementioned sensor difficulties (for NT2 students), it is not surprising to find significant differences in outcomes for the two groups. The specific differences were not unexpected: that NT1 learners have better vocabulary and poorer computer skills than NT2 learners is entirely in line with literature expectations (see [89, 156], which we used for participant selection in section 6.5.3). The findings reinforce once more that 'low-literate learners' are a very heterogeneous group [67], and that learning for these learners must be personalized to their wants and needs to be effective [78]. Furthermore, the found differences between male and female learners indicate that personalization can be valuable for many attributes. From earlier work results, we see particular participant attributes as 'less important': Age, sex, and schooling history did not show up as significant between-subjects factors in Deneka [217] or Massink [246]. The fact that learner sex is now a significant factor here indicates that specific scenario, types of learning support, and other factors of learning context can play a major role in determining what kinds of personalization are valuable. These results are comparable to Fourati et al. [273], who used an affectively expressive virtual storyteller character with children aged 6-10: The affective manipulation in that work had little main effects, but unexpected main and interaction effects showed difference between groups of children younger than 8 years and children aged 8 and above. Future work in this field should investigate along which personal attributes affective learning support for people of low literacy can best be personalized, using the results presented here as a starting-off point.

7

Discussion

”The world used to be black and white.
But now it’s a rainbow, full of colour.”

Leo, literacy ambassador

7.1. Conclusions

Societal participation is a serious bottleneck for low-literate citizens, who often lack the required practical information and communication skills, feelings of self-efficacy, and motivation. In this thesis, we proposed to develop a virtual learning environment (VLE) that could help people of low literacy individually train these skills in practical exercises and acquire self-efficacy and motivation. We called this environment VESSEL: *Virtual Environment to Support the Societal participation Education of Low-literates*. The thesis' general research question was:

How can VESSEL support low-literate people in achieving practical exercise success, self-efficacy, and motivation to participate?

This general research question was decomposed into five focused research questions and three hypotheses, following the Socio-Cognitive Engineering approach (SCE, cf. [84, 85, 163]).

7.1.1. Theoretical Foundation

The first step (Chapter 2) was focused on the collection of theory that could provide a relevant foundation for the VESSEL design (i.e. the operational demands, human factors knowledge, and technology) and on the creation of a first VESSEL specification, consisting of system objectives, functional requirements, and claims. This chapter's research question was:

7

Q1. Which operational demands, human factors knowledge, and technologies are important to the design of VESSEL, and which objectives, requirements, and claims can be derived from these?

Operational demands for VESSEL were drawn from the actor demographic of low-literate learners, the domain of societal participation behavior, and the task of learning. We used Kurvers et al.'s [89] five language learner profiles to delineate which low-literate learners we focused on. We excluded learners in the most advanced profile (1), who have high enough reading and writing skills and self-determination that learning support software will not help them much, and learners in the least advanced profile (5), who have too low reading and writing skill for a software solution to be applicable. Profiles 2, 3, and 4 describe low-literate learners from native Dutch (NT1) and non-native migrant (NT2) backgrounds, aged 16-65, with literacy levels between A1 and B1 [90]. We described societal participation as goal-directed social behaviour consisting of three elements: Language, societal knowledge, and participation skills, which are information and communication skills that are used in formal or informal social contexts. Building on Illeris [157], we defined that learning consists of cognitive, affective, and social processes, which should all be supported by VESSEL.

Human factors knowledge for VESSEL was taken from adult learning and ICT learning theory, and used to derive eight functional requirements for the VESSEL specification. We studied four adult learning theories: Andragogy, transformative learning, constructivism, and e-learning. From andragogy [123, 124], we derived requirement *R1. Adaptability*, expressing that all learners are different and that different learners require different approaches to learning and support. From transformative learning [126, 178, 274], we derived *R2. Sensitivity*, indicating that learning related to challenging topics requires a sensitive approach to be effective. From constructivism [23, 24, 128, 275], we derived *R3. Situatedness*, showing that learning is the active creation of knowledge, meaning learning should take place in the same context it will later be applied in. From e-learning [54, 65, 135, 136], we derived *R4. Collaboration*, describing learning as a collaborative effort that is improved by the presence of teachers and learning peers. Similarly, we studied four ICT learning principles: Information provision, worldwide communication, interactivity, and gaming principles. Information provision [56] suggested that ICT is inherently suited for displaying information in multiple modalities (leading to *R5. Multimodality*). Worldwide communication [55, 56] suggested that ICT allows learners to be in contact with teachers and peers at all times and regardless of barriers, meaning learning support can be provided whenever needed (*R6. Support*). Interactivity [25, 26, 56, 57] suggested that ICT can easily provide interactive learning options (*R7. Interactivity*). Finally, gaming principles [139–141] suggested that the application of digital gaming ideas and gamification could enhance learning effectiveness (*R8. Gaming principles*).

Two types of technology were studied for VESSEL. First, we assessed current practice by evaluating six examples of learning support software for low-literate learners. Five examples met between four and seven of our requirements, with the sixth meeting all eight. We found that all examples were intended as part of larger classroom learning packages, instead of as standalone solutions, and that relatively few incorporated skills training (3/6) compared to language learning (6/6) and societal knowledge (5/6). Since VESSEL is focused on skills training, self-efficacy, and motivation, it can coexist with existing software. Second, we investigated VLEs, software made up of digital environments, actors, and objects [38, 39], as a potential technology platform for VESSEL. We showed that VLEs are very well-suited to meeting all of VESSEL's requirements, except *R2. Sensitivity* (which almost exclusively covers learning content, regardless of technology).

We used the foundation to create an initial VESSEL design specification. The specification had one main system objective, which was high learning effectiveness. We incorporated the eight functional requirements, and defined between one and three claims for each requirement, with each claim explaining one way that this requirement contributed to high learning effectiveness. Each requirement could have one cognitive, one affective, and/or one social claim. Fig. 7.1 shows the specification for *R6. Support* as an example, including the learning effectiveness objective, requirement text, and claims.

7.1.2. Empirical Data

The second step (Chapter 3) aimed to expand the foundation with empirical data derived from the daily life experiences of people of low literacy (matching profiles

Objective: High learning effectiveness.

R6. Support: VESSEL should possess built-in support options.

- *Cognitive claim*: Built-in support options will aid learners in understanding the material, improving learning comprehension.
- *Affective claim*: Low-literate learners value the idea of being supported. Built-in support will set learners at ease while using the software, resulting in increased self-efficacy.
- *Social claim*: Built-in support options help learners feel supported. Learners are more motivated to continue if they know help is available, leading to higher learner retention.

Figure 7.1: Learning effectiveness objective, requirement text, and claims for R6. Support at the end of Chapter 2.

2-4) living in the Netherlands, and to use this data to validate and possibly refine the specification. This chapter's research question was:

Q2. How can we incorporate the subjective societal participation experiences of low-literates into the design of VESSEL, and which refinements to the VESSEL specification can be derived from this?

Including societal participation experiences into VESSEL's design required qualitative data acquisition suitable for low-literate participants, and subsequent data analysis. We chose to use participant workshops [15, 71] and cultural probes [72, 164, 166] to acquire rich unstructured data about the daily life experiences of low-literate people. We then used Grounded Theory [74, 168, 170–172] to transform these data into the *Societal Participation Experience of Low-Literates* (SPELL) model. This model provides an overview of personal, societal, and information-communication factors that are relevant to the societal participation experiences of people of low literacy, grounded in experiential data and ordered into hierarchical categories. It was incorporated into the foundation.

We used the SPELL model to validate and refine the system objectives, functional requirements, and claims in the VESSEL specification. The learning effectiveness objective was refined into sub-objectives: We defined that learning effectiveness has cognitive, affective, and social aspects, and that it is expressed in learning accessibility, learning experience, and learning outcomes, resulting in a 3-by-3 matrix of nine sub-objectives. The requirements were validated: All existing requirements were reflected in the data, and no new requirements were derived. The data were then used to refine the requirements by expanding on the core concepts. Finally, the claims were validated: All existing claims were reflected in the data, and between one and four new claims were derived for each requirement. Old and new claims were refined to make them explicitly refer to the nine sub-objectives. Fig. 7.2 shows the system objectives matrix, and Fig. 7.3 shows the refined specification for *R6. Support*.

	Learning Accessibility	Learning Experience	Learning Outcomes
Cognitive	*O1. Difficulty*: The learner can cognitive access and use the learning.	*O4. Performance*: The learner can follow the learning and interact with the material.	*O7. Success*: The learner reaches measurable learning goals.
Affective	*O2. Barriers*: The learner can emotionally engage with the learning.	*O5. Valence*: The learner can emotionally keep up with the learning.	*O8. Self-efficacy*: The learner's self-efficacy about societal participation increases.
Social	*O3. Intention*: The learner has the motivation and intention to start with the learning.	*O6. Engagement*: The learner has the motivation to keep up with the learning.	*O9. Retention*: The learner is motivated to come back to the learning later.

Figure 7.2: Matrix of learning effectiveness sub-objectives.

Objective: High learning effectiveness, as defined in sub-objectives O1 through O9.

R6. Support: 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.

- *Claim O1*: Learning support lowers the cognitive difficulty of learning.
- *Claim O2*: Learning support reduces affective barriers to learning.
- *Claim O3*: Learning support increases the intention of learners to start learning.
- *Claim O4*: Learning support results in better performance on exercises.
- *Claim O5*: Learning support improves emotional valence in the learning process.
- *Claim O6*: Learning support allows learners to engage with learning more.
- *Claim O7*: Learning support results in more successful exercise completion.
- *Claim O8*: Learning support results in higher self-efficacy for learners.
- *Claim O9*: Learning support increases learner retention.

Figure 7.3: System objectives, requirement text, and claims for R6. Support at the end of Chapter 3.

7.1.3. Proof-Of-Concept Prototype

The third step (Chapter 4) was to use the specification to create a proof-of-concept VESSEL prototype, consisting of four situated interactive exercises about information/communication skills and an Embodied Conversational Agent (ECA) coach that provided cognitive, affective, and social learning support for those exercises. This prototype was used to experimentally evaluate the effectiveness of the coach's learning support, focusing on *R6. Support* but incorporating all other requirements as well. We addressed two topics in this chapter: A design research question, and a hypothesis on learning effectiveness. The design research question was:

Q3. How can a VESSEL prototype with an ECA coach provide cognitive, affective, and social learning support that meets the operational demands and human factors knowledge?

To design this prototype, new operational demands, human factors, and technology were incorporated into the foundation. For operational demands, we defined that the exercises in VESSEL should be situated, interactive, scenario-based exercises, using topics from the list of crucial practical situations [2, 89]. Exercises should also address information or communication skills, and be set in formal or informal social contexts. Since the SPELL model shows that formal settings are seen as more challenging than informal settings, we chose the scenarios 'using online banking' (formal setting, information skills) and 'talking at a service desk' (formal setting, communication skills), and created one Easy and one Hard exercise for each, for a total of four exercises. In the Online Banking exercises, learners had to use a website to transfer money; in the Service Desk exercises, learners had to talk to an ECA character to report a lost passport.

7

For human factors, we incorporated theory to create cognitive, affective, and social learning support. Cognitive learning support based on (verbal) scaffolding [93, 204–206] was given *during* the exercises, to help learners understand and complete them. Affective learning support based on motivational interviewing [207, 211, 262] was given *after* the exercises, to help learners feel better about the learning process and increase their self-efficacy. Social learning support based on small talk [40, 153, 212] was given *before* the exercises, to establish a bond of liking and trust between the learner and the ECA and to improve learner motivation.

For technology, we created the ECA coach and the Service Desk exercise in the 'DRVret' Vizard environment from Brinkman et al. [58, 276]. A pre-study showed that our low-literate participants preferred the coach character to look like a young white woman in casual clothing, the Easy Service Desk ECA to look like a young white woman in formal dress, and the Hard Service Desk ECA to look like an unfriendly middle-aged white man [217]. All ECAs used prerecorded spoken voice lines in A1- to A2-level Dutch. The Easy and Hard Online Banking exercises were created in HTML.

The requirements in the VESSEL specification were refined to incorporate these choices. We chose to leave out requirement *R4. Collaboration*, which was the only one that necessitated the development of a multi-user prototype; use and

Objective: High learning effectiveness, as defined in sub-objectives O1 through O9.

R6. Support: 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.

- *Claims*: O1, O2, O3, O4, O5, O6, O7, O8, O9.
- **R6.1-C**: The coach should use verbal scaffolding techniques to offer cognitive learning support.
- **R6.2-C**: The coach should use motivational interviewing techniques to offer affective learning support.
- **R6.3-C**: The coach should use small talk to offer social learning support.

Use Case Service Desk: "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)"

Figure 7.4: Objectives, requirement text, sub-requirements, claims, and use case for R6. Support at the end of Chapter 4. Sub-requirement headers 'R6.1-C' through 'R6-3.C' indicate these sub-requirements describe coach functionality; No sub-requirements for exercise functionality are seen in this image.

evaluation of this requirement was instead deferred to later work. For each other requirement, we created sub-requirements for the coach and the exercises. Additionally, two use cases were written to describe the envisioned working of a VESSEL prototype. The system objectives and claims in the specification were not changed: We defined that claims for each original requirement also applied to its sub-requirements, necessitating no new claims. Fig. 7.4 shows the sub-requirements for *R6. Support*, and a small part of the use case 'Service Desk' that applies to this requirement.

A proof-of-concept VESSEL prototype was created, consisting of the four exercises and the ECA coach. The prototype was controlled via the Wizard-of-Oz method: Speech utterances for the coach and Service Desk exercise ECAs were activated by the Wizard operator, who used a simple ruleset to decide which utterance to use at what time. Because we wanted to evaluate the effectiveness of the ECA coach's learning support, two versions of the prototype were made: One with and one without the coach. An experiment was set up to compare the experience of using both versions: We predicted that the coach would increase learning effectiveness by improving six sub-objectives related to the learning experience and learning outcomes (Fig 7.2): *O4. Performance* (measured as learner performance during the exercise), *O5. Valence* (positive affect during the exercise), *O6. Engagement* (degree of user-system engagement), *O7. Success* (exercise completion rate), *O8. Self-efficacy* (self-efficacy regarding exercise topics and computer use), and *O9. Retention* (motivation to continue learning with VESSEL after the exercise). The main hypothesis for this experiment was:

H1. The VESSEL prototype with ECA coach that provides learning support results in a better learning experience and better learning outcomes than the VESSEL prototype without ECA coach.

In a mixed-method within- and between-subjects experiment, 12 participants used the VESSEL prototype with and without coach support in two experimental sessions. Of the six variables above, repeated-measures General Linear Model (GLM) analyses showed that learner performance, positive affect, and user-system interaction were higher when the coach was present, and that learner self-efficacy about the Hard Online Banking exercise increased after doing the exercise with coach support. These results support hypothesis H1. No significant results were found for exercise success, retention, and other categories of self-efficacy. Qualitative observations indicated that low-literate learners were both willing and able to engage with the coach. The offered learning support seemed particularly effective for and appreciated by the worst-performing learners.

7.1.4. Cognitive Learning Support

The fourth step (Chapter 5) was to formalize VESSEL's cognitive learning support. We created formal dialogue rules for the coach, describing how fast it should give support and what level of support it should give based on the learner's progress in the exercise, and defining how it should react to certain user speech utterances. We then experimentally tested if learning effectiveness could be improved by attuning the coach's support delay to individual learners' performance in previous exercises. We addressed a design research question and a learning effectiveness hypothesis. The design research question was:

7

Q4. How can a VESSEL prototype with an ECA coach provide formal rule-based cognitive learning support that meets the operational demands and human factors knowledge?

New operational demands, human factors, and technology were needed to design this prototype. Based on the proof-of-concept results, we incorporated three new online banking exercises into the operational demands, using the existing Hard Online Banking website. Each exercise consisted of eight critical steps: The user had to navigate to certain pages four times, and enter data correctly four times. We then made an overview of every potentially 'difficult element' that a learner could need cognitive support for: Every part of the website the user could interact with, and the actions needed for every critical step.

To create dialogue rules for cognitive support, additional literature on scaffolding was incorporated into the human factors [60, 204, 227, 230, 231, 233, 234, 236]. We defined five levels of cognitive learning support [204, 227, 230]: Prompts (the coach asks the user if they understand a concept or an exercise step), Explanations (the coach explains a concept or an exercise step), Hints (the coach provides oblique direction about what the user should do next), Instructions (the coach provides overt direction about what the user should do next), and Modeling (the coach demonstrates the correct next step to the user). One support utterance of each level was written for each difficult element, creating a corpus of support utterances. We defined that during an exercise, the coach knew which difficult element the user last interacted with (by clicking on it or entering data into it). If the user took no action for 20 seconds, the coach should give a support utterance for that element, starting at the lowest level (Prompt) and going up one level every time more support was needed for that element (to Explanation, then Hint, then Instruction, then Modeling). The time the coach waited between utterances was called the support delay: The coach could increase or decrease this delay based on the user's performance in their previous exercise. If the user needed Instruction or Modeling support for the majority of critical steps, the coach reduced the delay by 5 seconds to a minimum of 10. If the user needed Prompt or Explanation support for the majority of critical steps, or no support, the coach increased the delay by 5 to a maximum of 30. If the user primarily needed Hint support, no changes were made.

For technology, we formalized the coach's simulated speech recognition by defining a list of keywords it could detect, consisting of the name of every difficult element, "*yes*", "*no*", and the 'lack of understanding' category, containing all utterances indicating that the learner did not understand the coach, such as "*I don't understand*" and "*could you repeat what you just said*". We defined that the coach could understand user speech containing these keywords. If a keyword corresponding to a difficult element was used, the coach immediately used an Explanation support utterance for that element. If the coach has previously used an utterance in the form of a question, it could act on 'yes' and 'no' answers. For example, all Prompt utterances were written as yes-or-no questions, e.g. "*Do you know what online banking means?*". If a user answered 'no' to this, the coach would immediately give Explanation support for 'online banking', but if the user answered 'yes', the coach would skip the Explanation level next time it gave support and use the Hint level instead. Specific coach actions like this were written for all yes-or-no questions. Finally, if the user used a 'lack of understanding' utterance, the coach would immediately repeat its last-used utterance.

We incorporated the dialogue rules into the VESSEL specification, in the form of written rules and a control flowchart, along with a new use case. The requirements in the specification were also refined to incorporate the changes, as shown in Fig. 7.5. The objectives and claims were not changed.

A VESSEL prototype was created, consisting of the three new online banking exercises and the ECA coach that only used cognitive support rules. This prototype was controlled via the Wizard-of-Oz method: The Wizard operator tracked the difficult elements that the user interacted with, recognized relevant keywords in user speech, and used a stopwatch to track the support delay time. We used this prototype to evaluate if attuning the coach's support delay to user performance influenced learning effectiveness. The same sub-objectives from the previous study were used. We kept sub-objectives *O5. Valence*, *O6. Engagement*, and *O8. Self-efficacy* identical, but changed *O4. Performance* (required mental effort) and *O9. Retention* (learner's opinion of the coach's attitude and helpfulness), and measured

Objective: High learning effectiveness, as defined in sub-objectives O1 through O9.

R6. Support: 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.

- *Claims*: O1, O2, O3, O4, O5, O6, O7, O8, O9.
- **R6.1-C**: The coach should use dialogue rules based on verbal scaffolding techniques to offer cognitive learning support.
- **R6.2-C**: The coach should use motivational interviewing techniques to offer affective learning support.
- **R6.3-C**: The coach should use small talk to offer social learning support.

Use Case Online Banking: "The user tries to navigate to the correct page on the online banking website, but takes a long time doing so. After twenty seconds of the user not making any progress, the coach offers simple cognitive support. Example sentences: *"Do you know what you are supposed to do?" "Do you know what 'online banking' means?"* (R6.1-C, R7.1-C)"

Figure 7.5: Objectives, requirement text, sub-requirements, claims, and use case for R6. Support at the end of Chapter 5.

O7. Success in two ways: Required time to complete the exercise, and recall about exercise topics in a post-exercise test. We predicted that attuning the coach's support delay would improve *O4. Performance*, *O5. Affect*, *O7. Success* (lower required time and better recall) *O8. Self-efficacy*, and *O9. Retention*, but decrease *O6. Engagement*. The main hypothesis for this experiment was:

H2. The VESSEL prototype that attunes the support delay to learner performance between exercises results in a better learning experience and better learning outcomes than the VESSEL prototype that does not.

We carried out a mixed-method within- and between-subjects experiment with 28 participants doing three exercises over three weeks. For half of the participants, the coach attuned its support delay in exercises 2 and 3 based on their performance in exercises 1 and 2 respectively; for the other half, the coach always used a support delay of 20 seconds. Repeated-measures GLM analyses showed no significant differences in learning effectiveness between these two approaches. Hypothesis H2 was not supported. However, learner self-efficacy significantly increased after completing the first exercise in either experimental condition.

7.1.5. Affective And Social Learning Support

The fifth and final step (Chapter 6) was to formalize the coach's affective and social learning support, by creating dialogue rules for these support types and defining the necessary speech recognition keywords. We also incorporated a heart rate sensor and facial expression detection software, which were used to determine when to offer what kind of affective support. We then empirically evaluated if a

VESSEL prototype using cognitive, affective, and social learning support resulted in higher learning effectiveness than a prototype using only cognitive support. A design research question and a learning effectiveness hypothesis were addressed. The design research question was:

Q5. How can a VESSEL prototype with an ECA coach provide formal rule-based cognitive, affective, and social learning support that meets the operational demands and human factors knowledge?

New operational demands, human factors, and technology were added. We designed one new exercise, using a volunteer work scenario from the crucial practical situations. The Volunteer Work exercise had two components, related to information and communication skills: Learners filled out a form about their volunteer work preferences and experiences, and then talked to a 'recruiter' ECA character about volunteer work they wanted to do. This exercise was designed to offer cognitive, affective, and social challenges, and it was incorporated into the operational demands of the foundation.

Four new topics were incorporated into the human factors. First, we used emotion modeling theory [244, 247, 249, 252, 254, 257] to define that the coach recognized the negative valence emotions anger, sadness, and fear, and the positive valence emotion happiness. The coach offered affective support whenever the user experienced a negative emotion, but did not give support for positive emotions. We defined three levels of specificity for the support: General support (the coach understands that there is some affective issue, but cannot determine the exact emotion), Specific support (the coach knows the user's emotional state, but not the exact cause), and Very Specific support (the coach knows the user's emotional state and can link it to difficult elements in the exercise). Second, a review of motivational interviewing theory [207, 211, 261–263] was used to define four affective support levels: Reflective listening (the coach makes explicit what emotion it thinks the learner is experiencing), Normalizing (the coach attempts to put the learner's emotional reaction in a broader context), Affirmation (the coach affirms the learner's emotional reaction as a normal thing and helps them move forward), and Self-efficacy Supporting (the coach tries to raise the learner's self-efficacy). The following support utterances were created: One General utterance of each level, one Specific utterance of each level for each of the three negative emotions, and one Very Specific utterance of each level for each predefined emotional issue tied to a difficult element. Third, small talk theory [40, 41, 153, 200, 202, 212] was used to create a social support script, using three elements of small talk: Finding common ground, expressing interest in the conversation partner, and keeping the conversation at a non-threatening level of depth. The coach asked questions about the user's experience with volunteer work and told stories about its own experiences, used positive feedback to express interest in the user, and moved on from topics that the user did not want to talk about. Fourth, cognitive support utterances were created for the Volunteer Work exercise, using the approach described in Chapter 5.

Objective: High learning effectiveness, as defined in sub-objectives O1 through O9.

R6. Support: 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.

- *Claims*: O1, O2, O3, O4, O5, O6, O7, O8, O9.
- **R6.1-C**: The coach should use dialogue rules based on verbal scaffolding techniques to offer cognitive learning support.
- **R6.2-C**: The coach should use dialogue rules based on motivational interviewing to offer affective learning support.
- **R6.3-C**: The coach should use dialogue rules based on small talk to offer social learning support.

Use Case Volunteer Work: "The coach sees that the user's arousal goes up higher than normal. Since the coach sees an emotional reaction, but not clearly which emotion, it offers a general level of affective support. Example sentences: "*I can see that you are having a hard time.*" "*Many people struggle with this.*" (R1.1-C, R6.2-C, R7.1-C)"

Figure 7.6: Objectives, requirement text, sub-requirements, claims, and use case for R6. Support at the end of Chapter 6.

Two new technologies were incorporated. First, we used literature on autonomous emotion detection [255, 264–267] to choose two sensors the coach could use to detect the user's emotional state: The photoplethysmographic (PPG) heartrate sensor 'Shimmer' [269], which measured the user's heart rate and reported their derived arousal level, and the facial expression detection program 'FaceReader' [270], which measured the user's facial expression and used this to predict their emotional state and valence level. The coach gave affective support whenever the sensors showed that the user was experiencing an emotion. If FaceReader reported any negative emotion with over 50% certainty, Specific support was given for that emotion. If this happened when the user interacted with a difficult element that Very Specific support was defined for, this was given instead. If Shimmer reported high arousal, but FaceReader reported no emotion with high certainty, General support was given. For all cases, support was given immediately on detecting the emotion, or as soon as the coach was done with a previous support utterance. The four levels of affective support were given in the sequence of Reflective listening, Normalizing, Affirmation, and Self-efficacy Supporting. After giving all four utterances, the coach waited at least one minute before giving affective support again. And second, the ECA characters for this prototype were made in the Unity software platform. The ECA models available in Unity had a wider range of facial expressions and higher graphical fidelity compared to the Vizard models, making the models more affectively and socially expressive.

The affective dialogue rules and social support script were incorporated into the VESSEL specification, along with one new use case. The requirements in the specification were also refined to incorporate the changes, as shown in Fig. 7.6. The objectives and claims were not changed.

A VESSEL prototype was created consisting of the new Volunteer Work exercise and the ECA coach that used cognitive, affective, and social dialogue rules. This prototype was controlled via the Wizard-of-Oz method: The Wizard operator tracked the difficult elements that the user interacted with, recognized relevant keywords in user speech, used a stopwatch to track the support delay time for cognitive support, and used the Shimmer and FaceReader sensors to determine when to give affective support. We wanted to evaluate if giving cognitive, affective, and social support increased learning effectiveness over only giving cognitive support. Sub-objectives *O4. Performance*, *O5. Valence*, *O7. Success*, *O8. Self-efficacy*, and *O9. Retention* were kept identical to the previous study, while *O6. Engagement* was changed to measure the amount of user-coach interactions that users would start. We predicted that giving cognitive, affective, and social support would improve *O4. Performance*, *O5. Valence*, *O6. Engagement*, *O7. Success*, *O8. Self-efficacy*, and *O9. Retention* over only giving cognitive support. The main hypothesis for this experiment was:

H3. The VESSEL prototype that provides cognitive, affective, and social learning support results in a better learning experience and better learning outcomes than the VESSEL prototype that provides only cognitive learning support.

We performed a final mixed-method within- and between-subjects experiment with 34 participants. Participants did the Volunteer Work exercise twice in one experimental session, once while receiving only cognitive support and once while receiving 'full' (cognitive, affective, and social) support. Of the variables above, repeated-measures GLM analyses indicated that the Full Support condition resulted in higher social engagement between learner and coach. More learning support was given in the Full Support condition as well. These results partially support hypothesis H3. Results also showed that learner self-efficacy increased after completing the exercise the first time, regardless of experimental condition, and that learners required less cognitive learning support when doing the exercise a second time. Finally, both qualitative and quantitative results indicate significant differences in learning experience between learners in different sub-demographics: The experiences of learners in learner profiles 2, 3, and 4 were different from each other, and the experiences of men were different from those of women.

7.2. Scientific Contributions

The scientific aims of this thesis were to study aspects of the design, development, and evaluation of VESSEL. The scientific contributions derived from this are related to the data gathered in the foundation (new insights about learning and societal participation for people of low literacy, focused on information and communication skills), the VESSEL design specification (the effectiveness of the VLE, the ECA coach, and cognitive, affective, and social learning support), and the methods used in the evaluation (adaption of qualitative methods for research and requirements engineering with low-literate participants). These are described in turn.

We created a theoretical framework for low-literate learning by zooming in general learning theories (related to adult learning, ICT learning, scaffolding, motivational interviewing, and small talk) on the learning needs, goals, and limitations of people of low literacy, explaining the relevance and practical implications of each. Current work on low-literate learning mostly focuses on practical perspectives and results [32], or on the evaluation of specific technologies [52] or research and design methods [111]. Our framework represents a theory-driven view on the low-literate learning process, which is valuable for ensuring that requirements engineering and design with low-literate learners stays grounded in theory while still focusing on desired practical outcomes. This is expressed in our matrix of learning effectiveness: We combined Biggs & Moore's [161] and Cybinski & Selvanathan's [162] 3P-model of learning (*Presage*, *Process*, and *Product*) with Illeris' [157] cognitive, affective, and social aspects of learning in a 3-by-3 matrix describing learning effectiveness in terms of cognitive, affective, and social learning accessibility, learning experience, and learning outcomes, which we then zoomed in on low-literate learners to structurally describe what learning effectiveness for low-literate learners consists of.

The SPELL model is a theoretically and empirically grounded representation of the societal participation experiences of people of low literacy. We used the crucial practical situations [2, 3] and Schouten et al.'s [91] societal participation model of information and communication skills in formal and informal social contexts as a starting-off point to elicit participation experiences from low-literate participants, which were then used to create the SPELL model. The societal participation issues of (Dutch) low-literate citizens are often described from a third-person observer's perspective, focusing on the practical consequences of limited reading and writing skills, such as reduced schooling, unemployment, and low self-efficacy [1–4, 9, 10, 32, 89, 277]. The SPELL model incorporates daily life experiences to represent societal participation from the first-person perspective of low-literate citizens, showing cognitive, affective, and social issues from daily practice in equal measure. Examples include: The emotions low-literates experience in difficult situations, the times that they are and are not motivated to act and what happens when motivation is low, and which problems they see as their own fault or as the fault of governmental malice. Interestingly, while the SPELL model is (to our knowledge) the only model that describes the experiences of low-literates to this level of detail, it is not the only overview of barriers to participation in terms of personal, societal, and interaction factors: A recent study by Jokiaho et al. [278] lays out that the promises and benefits of e-learning approaches are held back by several barriers, which they divide into 'personal factors' (barriers related to the individual), 'institutional and cultural factors' (barriers related to the institution, outside of the individual's control), and 'technical factors' (barriers related to the implementation and use of technology). The SPELL model corresponds to the categorization of Jokiaho et al.'s overview.

Our prototype studies showed that all low-literate participants could operate VESSEL. There used to be a 'digital gap' between the digital skills of literate and low-literate people [87, 88], but this no longer seems to hold: Many low-literates own and use personal computers [68] and possess good computer skills [9]. Van Deursen & van Dijk [138, 279] define five categories of digital skills: Operational skills (being able to use the computer and Internet), formal skills (being able to navigate the Internet), information skills (being able to find and retrieve information online), communication skills (being able to communicate online), and strategic skills (being able to use computers and Internet for personal or professional gain). All VESSEL exercises require operational skills, and different exercises call for additional skills: The Online Banking exercises require formal and information skills (navigating the website and filling out the form), the Volunteer Work exercise requires information and communication skills (filling out the form and talking to the recruiter), and the Service Desk exercises require communication skills (talking to the employee) as well as what van Deursen & van Dijk [279] call intellectual skills (understanding how to interact with the government). The ECA coach supported formal, information, communication, and intellectual skills. Operational skills were not supported, but this turned out not to be a problem: All low-literate participants in profiles 2-4 possessed the operational skill level required for VESSEL. This further suggests that a digital gap for operational skills no longer exists. Different participants struggled with the other skill requirements for some exercises: NT1 participants had good communication skills, but poor formal and information skills, while NT2 participants had good formal skills, but poor communication skills (in Dutch).

To our knowledge, no studies have specifically investigated ECA use by people of low literacy. Our results have shown that low-literates were able to work with our ECA coach, and that the coach was beneficial to low-literate learners similar to what could be expected for literate learners: The coach improved the cognitive learning experience by supporting the learners with the exercise [48], the affective learning experience by making learners feel more positive, [198, 199] and the social learning experience by forming a bond with learners [201, 202]. This resulted in increased self-efficacy about the exercise topics [48]. Learners treated the coach as human, and appreciated its presence and support. However, this only works for low-literate learners if the coach is designed to match their cognitive, affective, and social limitations: The coach must use A1- to A2-level spoken language and express sensitivity about literacy issues.

We designed models of cognitive, affective, and social learning support for low-literate learners. Our cognitive support, based on scaffolding [204, 227, 230], was designed to ensure that all learners would eventually complete every exercise. This learning support was used as intended: Learners tried to complete exercises by themselves as much as possible, accepting support that was given but not relying on the coach to demonstrate everything. Our support is an example of 'light' scaffolding, which starts off with little support and increases over time as necessary. Most successful human tutors use 'heavy' scaffolding, which starts off with significant support and decreases this over time [236]. We chose to use light scaffolding as we were worried that heavy scaffolding would induce learners stop taking initiative and only follow the coach's instructions. These two approaches were not evaluated, but we speculate that heavy scaffolding would have been less effective:

Learners always followed Instruction-level support if the coach gave an utterance on that level.

Our affective support, based on motivational interviewing [207, 211, 261–263] and automated emotion detection output [269, 270], did not work as intended: No significant results were found. It is not clear if this means that our approach to affective support was flawed, or if using an ECA coach to give affective support to low-literate learners does not work. However, ECAs are capable of forming emotional bonds with learners [48, 280–283]. Lisetti et al. [280, 281] show that affective support based on motivational interviewing increased the acceptance of their ECA character. And individual effects were observed during studies: Some learners responded to affective support, and reported appreciating the coach's help. Consequently, we speculate that our particular approach to affective support was not effective, either because our motivational interviewing categories did not work as intended or because technical issues with the sensors hindered accurate use of the support (see section 7.4). Differences in affective support between the proof-of-concept and full prototypes could also be important: Affective support in the full prototype only addressed negative participant emotions and was given during the exercise, while affective support in the proof-of-concept prototype used participant successes and shortcomings as a jumping-off point for support after the exercise.

Our social support, based on small talk [40, 153, 200, 212], worked as intended. We saw interaction between learners and coach outside the parameters of the exercise: Learners would tell stories to the coach and ask it for its opinions. This shows that ECAs can be effective social actors with low-literate learners, if the ECA is designed with low-literate limitations in mind. Furthermore, we speculate that this effective social support could also have increased the effectiveness of the coach's cognitive support. Learners who received social support treated the coach as if it was human, including asking it for help without prompting and calling it by its name ('Anna'), while learners who received no social support more commonly treated the coach as an automated hint dispenser. Both de Rosis et al. [192] and Lane et al. [48] show that ECAs that socially bond with users achieve good cognitive and affective results, but do not evaluate if the social bonding actually causes this. More study is needed to see if low-literate learners are more likely to accept help from ECAs with human-like looks and characteristics.

Finally, we have demonstrated how participant workshops [71, 106] and cultural probes [72, 164, 166] can be used with low-literate participants. Our participant workshops incorporated directed exercises about crucial practical situations, which were used to illustrate examples and get participants thinking about their own experiences. Workshop leaders fostered an atmosphere of trust and sharing, which made participants willing to talk about their issues, and helped with the reading and writing elements in the exercises. Our cultural probes were designed to accommodate low-literate users: Included in the kit were a disposable camera, a voice recorder, and a set of pre-printed papers to write on. The cultural probes were introduced in one of the participant workshops, allowing us to explain the more complicated parts (how to use the voice recorder, and what to write down on the papers) and to get low-literate participants on-board. Iversen & Nielsen

[284] describe that cultural probes are valuable for studying hard-to-obtain real-life experiences. All 23 participants in our cultural probe studies returned highly personal recordings and pictures, showing that a well-designed probe can be valuable for reaching people of literacy. And we have shown how a shorter and more focused implementation of Grounded Theory [74, 167–170] can be used for rich data analysis and requirements engineering. Users of classical Grounded Theory warn that changing the method too far from its original design could erode its impact [172], while others argue that adapting the method to different circumstances makes it more accurate and widely applicable [171, 285]. We have demonstrated that a shortened Grounded Theory that omits theoretical sampling and the constant comparative method is valuable for requirements engineering, while still meeting Lincoln & Guba's [174] determinants for qualitative data correctness: Confirmability (results and claims are reflected in the data), credibility (results and claims are factually correct), transferability (results and claims are found in the real world), and dependability (results and claims could be repeated by other researchers).

7.3. Application Domain Contributions

At time of writing, 2.5 million people in the Netherlands are classified as low-literate [9, 17–19]. Our goal with VESSEL was to design a VLE that could complement existing classroom learning, allowing low-literate learners to train practical skills with learning support from the system whenever they want and increasing their self-efficacy about societal participation and their motivation to participate. The work in this thesis contributes to a better understanding of the participation experiences of people of low literacy, which could be valuable for people working in this domain, such as teachers, designers, researchers, and software developers.

The SPELL model visualizes how the societal participation experiences of low-literate people are influenced by personal traits, formal and informal societal factors, and attributes of information and communication tasks. For example, the model shows how low-literates may avoid starting a task if they fear it will be too difficult for them (personal trait), how they see the behaviour of large corporations as 'bullying' and stigmatizing low literacy (formal societal factor), and that for NT2 learners in particular phone calls are often a worse information medium than letters, because phone calls are quick and fleeting while letters are persistent (communication task attribute). The model is valuable for anyone who needs to take the experiences of low-literates into account: For instance, product designers could evaluate if low-literate people can use their designs (and are willing to do so), and government service providers could study what aspects of their behaviour come across negatively to low-literate users. The VESSEL design specification has similar value for people designing software or products for low-literate users, or related to low-literate learning: Driessen [286] has adapted the functional requirements in the specification to design a 'train-the-trainer' course for educators working with low-literate people.

In this work, we adapted the participant workshop and cultural probe methods for low-literate learners, and created our own quantitative measurement instruments based on visual analogue scales [243, 287, 288] and the Self-Assessment

Manikin [244]. Participant workshops and cultural probes are useful for domain practitioners who want to better understand the experiences of people of low literacy: Hanekamp [165] describes how teachers can use cultural probes to discover participation problems relevant to their students, which can then be used to design effective lessons. Our quantitative measures can be useful for engaging low-literates in research or design work. Munteanu et al. [289] describe that this is often difficult, because traditional data collection measures generally do not take low literacy into account. Our measures were specifically designed for low-literate users, and they were generally understood well and used correctly (but see section 7.4).

Finally, qualitative observations from the studies showed that almost all participants genuinely wanted to help us study VESSEL. Participants were interested in the novelty of the digital coach, and the possibility of getting access to computer support. We initially worried that the technology would be too complex for low-literate users to engage with, or that the users would not be interested in talking to a digital character, but this proved unfounded. Debriefing showed that participants really appreciated how VESSEL was designed with their needs and limitations in mind, including an appropriate language level, a calm and friendly coach character, and exercises that could be done with either simple mouse and keyboard actions or by talking to the computer. Additionally, we worked extensively with local teachers to introduce ourselves to participants beforehand, and to position the work not as 'scientific study into the effectiveness of software design', but as 'these researchers need your help to make this software for low-literate people as good as they can'. While we do not know how participants would have behaved if no teachers or researchers had been around, people who chose to participate in our experiments were generally interested in the system and the work, and committed to completing every session. Even in studies with multiple sessions across several weeks (Chapters 4 and 5), no participants dropped out.

7.4. Limitations

Six types of limitation were identified in this work. First, our three experimental studies had relatively low participant numbers: 34 participants completed the third experiment (Chapter 6), 28 participants completed the second experiment (Chapter 5), and only 12 participants completed the first experiment (Chapter 4). This calls the statistical power of our results into question. Furthermore, these participants groups were not fully representative samples of Dutch low-literate language learners. We tried to ensure representativeness by using Kurvers et al.' [89] language learner profiles to select low-literate language learners from profiles 2 (high-skilled NT1 and NT2 learners), 3 (medium-skilled NT2 learners), and 4 (medium-skilled NT1 learners). However, in practice, learners from profile 2 were overrepresented compared to profiles 3 and 4. The reason for this limitation is that we decided to recruit our participants and conduct our experiments on-site at low-literate language classes in the Netherlands (*e.g. in Den Helder, Utrecht, Nijmegen, Wijchen, Beuningen, Groesbeek, and Delft*). We made this decision both to make our experiments more accessible to low-literate participants (both the 'literacy ambassadors', see Chapter 1, and the language teachers we consulted strongly encouraged this approach), and to ensure ecological validity of our results. Although our participant numbers are somewhat low for statistical analysis, it is worth noting that we managed to reach a relatively large section of Dutch language learning classes, and that we recruited an impressive number of low-literate language learners to participate in our experimental evaluation of a virtual learning environment.

Second, related to the above, our focus on language learning classes restricted our recruitment to only one particular type of low-literate Dutch citizen, leaving out other types. Individual differences notwithstanding, all low-literates found in language classes have the means, the ability, and the desire to be in a language class. This does not hold true for all low-literates. For instance, Van der Meer et al. [290] describe a persona-based categorization of low-literates based on their life experience and view of education, ranging from the 'focused improver' (*doelgerichte kansverbeteraar*), who only needs help finding the best language class for their situation, to the 'innocent rascal' (*onschuldige kwajongen*), who will never improve on their own, requiring pressure from a trusted authority figure to interact with language learning. Consequently, it is unclear how well our results will generalize outside language learning classrooms.

Third, the prototypes in Chapters 4 and 5, which were built on the Vizard framework, suffered from lag when playing support utterances: Utterances were not played directly after selecting, but between two and twenty seconds later. In Chapter 4, the delays were highly variable, making the delivery of cognitive support uncertain. In Chapter 5, the delays were always 5 seconds, meaning the wizard could manage this to a degree. But noise was still present: If participants asked new questions or took new actions in the 5 seconds between selecting an utterance and that utterance being played, the wizard had to either let the old utterance play, or mute the system in preparation for a new utterance. As support delay was an experimental manipulation in this chapter, the noise of this input lag has likely lowered the strength of the results.

Fourth, all dialogue rules and speech recognition for the prototypes in Chapters 4 to 6 were controlled using the Wizard-of-Oz method. Consequently, we cannot know if a fully autonomous VESSEL would achieve the same results. Wizards followed the dialogue rules as closely as possible, but had to make snap decisions in each study. In Chapter 4, rules for cognitive support and speech recognition during the exercise were not detailed enough for the wizards to fully rely on. In Chapter 5, wizards had to react to occurrences of input lag, as described above. In Chapter 6, wizards had to choose what affective support to provide if the learner expressed emotions that the sensors did not pick up (for instance, expressing emotion in body language, or looking away from the camera while angry). And in all studies, the wizards emulated a level of speech recognition that current technology might not be able to replicate. The effectiveness of speech recognition for non-native speakers in particular is generally poor [291]: Beginning NT2 learners, who speak Dutch haltingly and with strong accents, may not reach the same level of effectiveness with VESSEL if fully autonomous speech recognition were used.

Fifth, technical issues were encountered with the Shimmer and FaceReader sensors used in Chapter 6. Both sensors were markedly less effective on participants of colour. This issue is not unique to our work: Many studies on facial recognition only use white participants, or employ algorithms optimized for white skin [292, 293]. In our work, the bright lighting of the experimental locations may have introduced additional noise [294]. Our pilot testing for this study was not sufficient: We only tested white participants in a controlled environment. Accurate pilot testing, with all participant groups and under experimental circumstances, would have identified this problem earlier. Since we used autonomous emotion detection to avoid human bias in identifying participant emotion and ensuring that standardized affective support would be given to all participants, these outcomes also raise the question of whether or not this approach is generally feasible.

Finally, the Likert scale questionnaire used in Chapter 4 and the Visual Analogue Scale questionnaires used in Chapters 5 and 6 were developed by us for the purposes of this work, and were not externally validated. As few existing measurement tools were designed with people of low literacy in mind, we developed our own, based on knowledge of and experience with the low-literate learner demographic. Qualitative observations seem to suggest that participants understood our measures and used them correctly, but lack of validation remains an issue. Additionally, different measures were used in Chapter 4 (Likert scale) and Chapters 5 and 6 (Visual Analogue Scale), making the results from these studies harder to compare.

7.5. Directions For Future Work

We suggest five directions for future work. First, the experimental studies in Chapters 4 to 6 should be replicated with larger participant populations divided equally over language learner profiles 2, 3, and 4. This addresses the previous section's first limitation. Additionally, prototypes for these studies should be developed to no longer rely on the Wizard-of-Oz method, and (ideally) no longer suffer the technical issues we reported with regard to support utterance input lag and the accuracy of the Shimmer and FaceReader sensors, addressing the second, third, and fourth limitations. For all studies, we suspect that skewed participant distribution and technical issues have lowered the strength of the findings. Repeating these studies under optimal circumstances can show whether or not this was the case. Replication work should also either externally validate our measures, or adapt existing validated measures to low-literate learners, addressing the final limitation.

Second, not all parts of the VESSEL design specification were studied: The requirement *R4. Collaboration* was left out of every prototype design, as we focused on evaluating single-user prototypes. However, theory affords high importance to collaborative learning, and the SPELL model shows that low-literate learners greatly value the presence of peers and teachers. Future work should develop and evaluate multi-user VESSEL prototypes. This could take the form of multiple learners doing one exercise on the same computer, VESSEL being integrated into classroom learning, or learners doing exercises together over the Internet. Similarly, we have not applied the Gaming Principles requirement very broadly. The use of digital gaming and gaming principles in educational spaces holds promise for improving learning effectiveness: For instance, Antonaci et al. [295] describe how gamification could be used to address the gap between learner intentions and learner outcomes in Massive Open Online Courses. Future work should investigate if incorporating gaming principles throughout the design of VESSEL will improve the learning experience, as well as study how these gaming principles should be adapted to be valuable for low-literate learners.

Third, our work has repeatedly looked at the learning experience of working with VESSEL and its effect on learning outcomes, but has not investigated VESSEL's accessibility, as our experimental designs were not suited for testing this. Since we consider learning accessibility an important area of learning effectiveness, future work should investigate if VESSEL actually contributes here. While cognitive accessibility refers to learners' ability to work with the system, affective and social accessibility refer to learners' willingness and motivation to use the system. These areas could be tested using free-choice experiments, wherein participants are free to spend as much or as little time with the system as they want [296, 297], or in the form of experiments where participants use VESSEL at home, without supervision, for a longer period of time (cf. [13, 298]).

Fourth, our studies focused on the short-term learning effectiveness impact of VESSEL. This raises two issues. On the one hand, the study in Chapter 5 (which attuned the coach's support delay to learner performance) found no results in three weeks of study. It is possible that the envisioned benefits of this adaptation take longer to become apparent, which calls for longer future experiments. On the other hand, results from all experiments showed that learner self-efficacy increased after the first exercise and remained high throughout the 1-3 weeks of study. Future work should investigate whether or not this higher self-efficacy remains stable over longer periods, and whether or not it depends on learners regularly practicing with VESSEL. In general, our studies did not investigate VESSEL's long-term impact on the societal participation of people of low literacy. While our experiments showed that low-literate learners benefit from VESSEL in the short term, future work should study if these effects persist in the long run, improving low-literate societal participation as a result.

Finally, in the introduction to this work, we suggested that VESSEL can address shortcomings to classroom learning by making learning accessible to more learners, making it easier to individualize learning, and allowing learners to practice scenarios on the computer that are difficult to practice in the classroom. Future work should investigate in what ways VESSEL can be integrated into the existing classroom learning ecosystem, aiming to combine the advantages of both approaches to maximize learning effectiveness for low-literate learners. This could take the form of *blended* learning, which mixes traditional classroom work with digital or e-learning [65]: VESSEL seems to lend itself to this sort of approach, and both Steehouder & Tijssen [67] and Driessen et al. [66] have already demonstrated that blended learning can be effective for low-literate learners, by increasing the accessibility and applicability of learning materials and complementing the role and responsibilities of language teachers with ICT. However, Jokiaho et al. [278, see section 7.2] clearly outline the potential barriers and accessibility risks associated with e-learning. As low-literate learners are particularly vulnerable to any barriers to accessibility, follow-up work should take extra care to consider whether or not any broad-scale implementation of e-learning will be effective at reaching this demographic.

7.6. Final Message

One valuable result that should not be ignored is this: The low-literate people who participated in our work *liked* VESSEL and the coach. When introducing the work beforehand, and when debriefing participants afterwards, the mood was almost always one of interest and excited optimism. A big problem surrounding low literacy in general is that people of low literacy feel that they aren't being listened to, and that their problems are downplayed or ignored by literate people who do not empathize. In contrast, the core tenet behind VESSEL and our user-centered design approach was that we were very explicitly listening to the needs, wishes, and practical experiences of people of low literacy. People enjoyed working with VESSEL because VESSEL was making it clear that it took their perspectives into account. If we can keep this up, and implement VESSEL such that it consistently makes people of low literacy feel valued, included, and taken seriously, that alone will be a very valuable result.

To learn more about the COMMIT/'Social Conventions Learning in Mixed Reality' project that formed the work described in this thesis, and to see a demonstration of the third VESSEL prototype carried out by real low-literate participants (including literacy ambassador Leo), visit: https://www.youtube.com/watch?v=vOZjPP_NO1k.

Acknowledgements

The creation of the thesis you hold in your hands was supported by a large number of people. Let's see if I remember to thank them all.

First and foremost, obviously, I want to thank my promotor, Mark Neerincx, and my copromotor and daily supervisor, Anita Cremers. Thank you for the four years of teaching, support, and collaboration while we worked on our projects together, and the additional four years of support while I slowly clawed my way over the finish line. You taught me both the theory and the practice of what it means to be a good scientist, and while you were always patient with me whenever things took time, you never stopped encouraging me to reach for the stars. Mark, I to this day remain impressed by the amount of expertise and specialized knowledge you could bring to bear seemingly without trying, and the patience with which you walked me through your ideas until I really, fully understood them. Anita, if there's one thing I'm grateful for above anything else, it's that you always made time to support me and work together with me when needed. You always treated me as an equal member of the research team, and that made me feel welcome and valued from day one. This thesis would not exist without the both of you in more ways than the obvious.

On the other side of the coin, I want to thank the four students I supervised and collaborated with on various parts of this project: Isabel Pfab, Oanez Deneka, Pim Massink, and Fleur Venneker. Isabel, you came into our project with a tremendous energy and clear ideas about the research you wanted to do, and it's no exaggeration to say you had a lasting impact during your short stay with us – it took them *so long* to take your nameplate down. While the work we did didn't make it into this thesis, it *did* take us to Paris, and getting lost on Île Saint-Louis at night is a story I still tell to this day. Oanez, your self-directed drive and work ethic were so impressive that I still wonder if you really needed my 'supervision' at all. As one of the main proponents of the idea of an ECA coach, the work you did (Chapter 4) incontrovertibly changed the shape of our project and of this thesis. It was a joy to work with you, and I hope New Zealand is treating you as well as you deserve. Pim, you were never afraid to tell me when my supervision strategies were not giving you what you needed, and I consider myself a better supervisor through what I learned in our interactions. You were an unstoppable reader, thinker, planner, and modeler: Anyone who knows you can easily recognize your signature in Chapter 5's extensive theory work. And while my family and friends might think that my newfound appreciation for whiteboards came out of the blue, you and I know better. Fleur, coming in on the tail end of a four-year project shaped by a dozen people, you did an incredibly admirable job of implementing your own ideas and putting your own spin on the work we did (Chapter 6). I'll never forget coming back from a week of vacation to find that you completely re-imagined the experimental context in my absence, nor will I forget realizing you were right. Of the three experiments I did with the coach, our version had the most rules and moving parts, but you made it look easy. Thank you, all of you, for making this thesis into what it is. By extension, I also want to thank each of your university supervisors: Betsy van Dijk, Mariët Theune, Stella Donker, and Tibor Bosse.

I want to thank the members of the thesis reading committee. I am very grateful for the valuable time they spent reading and evaluating this thesis, and the comments they provided have pushed this thesis to become better than it was. Thank you for choosing to be a part of this.

The project that this thesis was a part of arose from collaboration between TNO Soesterberg and TU Delft. As such, I was lucky to have two excellent work places in these four years, and two great sets of colleagues that supported me, spent time with me, pushed me to be better, and made this part of my life a time I will always remember fondly. Of the two, I spent most of my time at ever-scenic TNO Soesterberg, which very quickly became my home away from home; I remember afternoon forest walks, eating egg rolls from the market food truck, Christmas celebrations, and the occasional ping-pong tournament. I want to thank everyone who was a part of that experience: Angelique, Anne-Marie, Arnold, Bruno, Conchita, Constance, Ellen, Hanna, Iris, Jacomien, Joris, Jurriaan, Kim, Roald, Rosemarijn, Ruth, Tina, Vanessa, Willie, and anyone and everyone else. I want to extend special thanks to three people in particular: Judith Kessens, for expertly keeping the project together in the early years, Nanja Smets, for supporting me from literal day one when I was dropped in your office without warning, and Rosie Paulissen, both for helping shape this project in general and for lending your voice to the coach over and over, shaping this project in more ways than one. Delft, in contrast, was always more of a working vacation, a necessary and at times welcome break from my routine that came with its own experiences; here I remember playing board games after work hours, walking to the sports center, and repeatedly trying (and only sometimes failing) to find each other in the over-full first-floor EWI cafeteria. I want to thank: Arman, Bart, Catholijn, Chang, Chao, Chris, Christina, Corine, Dwi, Hani, Junchao, Judith, Madalina, Myrthe, Nelie-Roos, Ni, Pietro, Reyhan, Ruud, Tim, Timi, Tingting, Tjerk, Ursula, Wenxin, Willem-Paul, Yun, and anyone and everyone else. I want to extend special thanks to Alex Kayal, who I feel was in the same boat as me more than anyone else: Same project, same year, same Graduate School, same difficulty getting the people we needed. To this day I can't see Hendrik-Ido-Ambacht without smiling. By extension, I also thank to thank the COMMIT project for making all of the above possible.

I want to extend further thanks for everyone else I worked with on this project. Marianne Driessen, Marieke Hanekamp, Annemarie Groot, and Knup Fuhri from CINOP, for making sure that VESSEL was as good and as comprehensible as possible. The language class teachers we worked with, including Marysia Bocharewicz, Will Kocken, Conny Brouwer, Mia Pouwels, Gertie Brandts, Jet Wijdeveld, Ivonne Brand-Verhagen, and everyone from Stichting ABC, for putting us in touch with low-literate language learners willing to work with us. And every single one of those language learners for putting your trust in us and helping us test our design. Your feedback was always invaluable for staying on the right path.

Of course, I owe a huge debt of gratitude to all my friends and family. I'm grateful to my parents, Ineke and Peter, my brothers, Glenn and Evan, and my sister-in-law, Dagmar, for supporting me and putting up with my nonsense when this PhD took a little more time than usual, then took a little more time again. I'm grateful to my extended family for their ongoing love and support: Petra, Gerard, Hans, Twan, Marc, Susanne, Ellen, Marco, Ellie, Ab, Annabel, Jan, Mats, Aimée, Ellenore, Sophie, Stijn, Famke, and Derek. And most recently Milo Dylan Schouten and Jason Evan Schouten, who will one day learn what a source of inspiration they've been. I'm grateful to all my friends in the Netherlands, who kept ensuring I maintained a healthy work-life balance (in both ways): Koen, Jasper, Miel, Marco, Bram, Bart, Richard, Simon, Maurice, Andrew, Roos, Sanne, Eva, Shirley, and Izzy. And I'm grateful to all of my online friends, the world over: Jasmine, Andrew, Zach, Ariel, Carey, Arron, Loki, Rita, Josh, Jacob, Melissa, Sheila, Matthias, David, Alan, Greg, Silas, Shoshanna, Jeffry, Brandon, Brandon, Alex, Muse, Maddy, Arthur, Taylor, and everyone else I may be forgetting. All of you have no idea to what degree you've kept me sane throughout this whole ordeal.

And last, but definitely not least, Rajeshwari Manojna Adavi. My Manu. I can almost hear you protest as I write this: That I only met you after most of the writing was over, that everyone else in these acknowledgements deserves the accolades more than you do, that you had no real influence on this thesis becoming what is it. But when the finish line seemed like a distant impossibility and it felt like the weight of everything was bearing down on me all at once, it was your encouragement and your support and your love that gave me the strength to keep going. You inspire me to be better than I am, baba, and I am grateful beyond words for every day you choose to be a part of my life.

Bibliography

References

[1] M. de Greef, *Leren voor leven: een eigen plek in het dagelijks leven*, Tech. Rep. (Spectrum GMO Gelderland, Velp, 2009).

[2] J. Kurvers and I. van de Craats, *Literacy and Second Language in the Low Countries,* in *Low-Educated Second Language and Literacy Acquisition Proceedings of the Third Annual Forum*, edited by M. Young-Scholten (Durham: Roundtuit Publishing, Newcastle upon Tyne, England, 2007) pp. 17–23.

[3] I. van de Craats, *Obstacles on highway L2,* in *Low-educated second language and literacy acquisition: Research, policy and practice* (Radboud University Nijmegen, 2007) pp. 149–163.

[4] M. Buisman and W. Houtkoop, *Laaggeletterdheid in kaart*, Tech. Rep. (Expertisecentrum Beroepsonderwijs & Stichting Lezen en Schrijven, 2014).

[5] Y. Eshet-alkalai, *Digital literacy: A conceptual framework for survival skills in the digital era,* Journal of Educational Multimedia and Hypermedia **13**, 93 (2004).

[6] S. B. Merriam, *Adult Learning,* in *Adult Education - Adult Learning, Instruction and Program Planning* (Elsevier Ltd, 2010) pp. 12–17.

[7] G. van Oenen, *Liberaal denken, informaliteit en menselijke maat,* Migrantenstudies **18**, 93 (2002).

[8] T. S. Murray, I. S. Kirsch, and L. B. Jenkins, *Adult Literacy in OECD Countries: Technical Report on the First International Adult Literacy Survey*, Tech. Rep. (U.S. Department of Education, Office of Educational Research and Improvement, 1997).

[9] M. de Greef, M. Segers, and J. Nijhuis, *Feiten & Cijfers geletterdheid*, Tech. Rep. (Stiching Lezen & Schrijven, 2014).

[10] M. de Greef, M. Segers, J. Nijhuis, and J. F. Lam, *Impactonderzoek taaltrajecten Taal voor het Leven door Stichting Lezen & Schrijven op het gebied van sociale inclusie en leesvaardigheid*, Tech. Rep. (Maastricht University, Maastricht, 2014).

[11] H. Mertens and R. van het Zwet, *Het versterken van de maatschappelijke participatie van oudere migranten*, Tech. Rep. (Movisie, Utrecht, 2009).

[12] COMMIT, *COMMIT: A Public-Private ICT Research Community,* (2012).

[13] A. Kayal, *Normative Social Applications*, Ph.D. thesis, Delft University of Technology (2017).

[14] COMMIT, *IUALL: Interaction For Universal Access,* (2012).

[15] E. S. Wilschut, J. G. M. de Jong, R. van Son, and A. H. M. Cremers, *ICT-ondersteuning voor ouderen ter bevordering van sociale veiligheid in de wijk,* Tijdschrift voor Veiligheid (2014).

[16] OECD, *Literacy in the Information Age: Final Report of the International Adult Literacy Survey*, Tech. Rep. (Statistics Canada, 2000).

[17] A. Groot, K. Coppens, and J. F. Lam, *Motiveren van laaggeletterden: Een literatuurstudie naar succesvolle interventies.* ECBO (2019).

[18] Algemene Rekenkamer, *Aanpak van laaggeletterdheid*, Tech. Rep. (Algemene Rekenkamer, 2016).

[19] M. de Greef, M. Segers, and J. Nijhuis, *Feiten & Cijfers Laaggeletterdheid: De invloed van lage basisvaardigheden op deelname aan de maatschappij*, Tech. Rep. (Maastricht University, Maastricht, 2018).

[20] M. Broens, *Werven met nuance: Klasse!, een kans voor de aanpak nt1,* Sociaal Bestek **81**, 16 (2019).

[21] J. Gubbels, A. M. L. van Langen, N. A. M. Maassen, and M. R. M. Meelissen, *Resultaten PISA-2018 in vogelvlucht*, Tech. Rep. (Universiteit Twente, Enschede, 2019).

[22] M. M. M. Peeters, *Personalized Educational Games - Developing Agent-supported Scenario-based Training*, Ph.D. thesis (2014).

[23] P. L. Berger and T. Luckmann, *The Social Construction of Reality: A Treatise in the Sociology of Knowledge* (1966).

[24] D. H. Jonassen, *Objectivism versus Constructivism: Do We Need a New Philosophical Paradigm?* Educational Technology Research and Development **39**, 5 (1991).

[25] D. A. Kolb, *Experiential learning: experience as the source of learning and development*, 1984 (Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1984).

[26] D. A. Kolb, R. E. Boyatzis, and C. Mainemelis, *Experiental Learning Theory: Previous Research and New Directions,* in *Perspectives on cognitive, learning, and thinking styles*, 216, edited by R. J. Sternberg and L. F. Zhang (NJ: Lawrence Erlbaum, 2000).

[27] M. Gijsberts, W. Hujink, and J. Dagevos, *Jaarrapport integratie 2011*, Tech. Rep. (Sociaal en Cultureel Planbureau, Den Haag, 2012).

[28] R. van der Aa, S. Koppert, S. van Geel, M. Buisman, J. Meijer, and H. Sligte, *Evaluatie Actieplan Laaggeletterdheid 2012-2015*, Tech. Rep. (Ministerie van Onderwijs, Cultuur, en Wetenschap, Rotterdam, 2015).

[29] P. Steehouder and P. Baay, *Toekomst van de educatie vanuit het perspectief van ROC's*, Tech. Rep. (ECBO, Stichting Lezen & Schrijven, 's Hertogenbosch, 2016).

[30] T. Bersee, *Veertig jaar aanpak laaggeletterdheid: ambitie, onvermogen en lange adem,* (2019).

[31] Sociaal-Economische Raad, *Samen werken aan taal: Een advies over laaggeletterdheid*, Tech. Rep. April (Sociaal-Economische Raad, Den Haag, 2019).

[32] M. de Greef, *Leren in verschillende contexten; De opbrengsten van trajecten taal- en basisvaardigheden voor laagopgeleiden*, Tech. Rep. (Arteduc/CINOP, 's Hertogenbosch, 2012).

[33] R. Bijl and A. Verweij, *Monitoring integration in the Netherlands,* in *Measuring and monitoring immigrant integration in Europe*, edited by R. Bijl and A. Verweij (Sociaal en Cultureel Planbureau, Den Haag, 2012) Chap. 12, pp. 239–252.

[34] P. Shea, E. Fredericksen, A. Pickett, W. Pelz, and K. Swan, *Measures of Learning Effectiveness In the SUNY Learning Network,* Online Education **2**, 31 (2001).

[35] V. J. Shute, E. A. Graf, and E. G. Hansen, *ETS Research Report Series*, Tech. Rep. 01 (2006).

[36] V. J. Shute, E. G. Hansen, and R. G. Almond, *An Assessment for Learning System Called ACED: Designing for Learning Effectiveness and Accessibility,* ETS Research Report Series **2007**, i (2007).

[37] V. J. Shute and D. Zapata-Rivera, *Adaptive educational systems,* Adaptive technologies for training and education **7** (2012).

[38] J. Bishop, *The role of mediating artefacts in the design of persuasive e-learning systems,* Proceedings of the First International Conference on Internet Technologies and Applications (ITA 05) , 54 (2005).

[39] H.-M. Huang, U. Rauch, and S.-S. Liaw, *Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach,* Computers & Education **55**, 1171 (2010).

[40] T. W. Bickmore and J. Cassell, in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 396–402.

[41] J. Cassell, J. Sullivan, E. Churchill, and S. Prevost, *Embodied Conversational Agents* (MIT Press, 2000).

[42] R. Moreno and T. Flowerday, *Students' choice of animated pedagogical agents in science learning: A test of the similarity-attraction hypothesis on gender and ethnicity,* Contemporary Educational Psychology **31**, 186 (2006).

[43] C. Nass, K. Isbister, and E.-J. Lee, *Truth is beauty: Researching embodied conversational agents,* Embodied conversational agents , 374 (2000), arXiv:arXiv:1011.1669v3 .

[44] Y. Kim and A. L. Baylor, *Pedagogical Agents as Social Models to Influence Learner Attitudes,* Educational Technology **47**, 23 (2007).

[45] C. Nass, J. Steuer, and E. R. Tauber, *Computers are social actors,* Conference companion on Human factors in computing systems - CHI '94 , 204 (1994).

[46] B. Reeves and C. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places* (Cambridge University Press, New York, 1996).

[47] P. Skalski and R. Tamborini, *The role of social presence in interactive agent-based persuasion,* Media Psychology **10**, 385 (2007).

[48] H. C. Lane, C. Cahill, S. Foutz, D. Auerbach, D. Noren, C. Lussenhop, and W. Swartout, *The effects of a pedagogical agent for informal science education on learner behaviors and self-efficacy,* in *Artificial Intelligence in Education: 16th International Conference*, edited by K. Yacef, H. C. Lane, J. Mostow, and P. Pavlik (2013) pp. 309–318.

[49] T. W. Bickmore, L. Pfeifer, and B. W. Jack, *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* (2009) pp. 1265–1274.

[50] D. Schulman, T. W. Bickmore, and C. L. Sidner, *An Intelligent Conversational Agent for Promoting Long-Term Health Behavior Change using Motivational Interviewing,* in *AI and Health Communication - Papers from the AAAI Spring Symposium (SS-11-01)* (2011) pp. 61–64.

[51] M. Macedonia, I. Groher, and F. Roithmayr, *Intelligent virtual agents as language trainers facilitate multilingualism,* Frontiers in Psychology **5**, 1 (2014).

[52] P. Wik, *The Virtual Language Teacher: Models and applications for language learning using embodied conversational agents*, Ph.D. thesis, KTH School of Computer Science and Communication (2011).

[53] P. Wik and A. Hjalmarsson, *Embodied conversational agents in computer assisted language learning,* Speech Communication **51**, 1024 (2009).

[54] D. R. Garrison, *E-Learning in the 21st Century: A Framework for Research and Practice*, 1st ed. (Routledge, New York, New York, USA, 2011).

[55] K. B. Nielson, *Self-study with language learning software in the workplace: What happens?* Language Learning & Technology **15**, 110 (2011).

[56] C. Richards, *The Design of Effective ICT-Supported Learning Activities: Exemplary Models, Changing Requirements, and New Possibilities,* Language Learning & Technology **9**, 60 (2005).

[57] M. Barak, *Instructional principles for fostering learning with ICT: teachers' perspectives as learners and instructors,* Education and information technologies **11**, 121 (2006).

[58] W.-P. Brinkman, C. van der Mast, G. Sandino, L. T. Gunawan, and P. M. G. Emmelkamp, *The therapist user interface of a virtual reality exposure therapy system in the treatment of fear of flying,* Interacting with Computers **22**, 299 (2010).

[59] F. Cornelissen, M. A. Neerincx, N. J. J. M. Smets, L. Breebaart, P. Dujardin, and M. Wolff, *Gamification for Astronaut Training,* in *The 12th international conference on space operations, Stockholm, Sweden* (2012).

[60] R. J. Mislevy, A. Oranje, M. I. Bauer, A. von Davier, J. Hao, S. Corrigan, and E. Hoffman, *Psychometric Considerations in Game-based Assessment* (GlassLab, 2014).

[61] J. Linssen and M. Theune, *Using Out-of-Character Reasoning to Combine Storytelling and Education in a Serious Game,* in *Proceedings of the 1st International Workshop on Pedagogically-driven Serious Games (PDSG 2012)* (2012) p. 4.

[62] J. Linssen, M. Theune, M. Bruijnes, and T. de Groot, *Beyond Simulations: Serious Games for Training Interpersonal Skills in Law Enforcement,* (2014).

[63] J. Linssen, M. Theune, and D. Heylen, *Taking Things at Face Value : How Stance Informs Politeness of Virtual Agents,* (2012).

[64] R. Blake, N. L. Wilson, M. Cetto, and C. Pardo-Ballester, *Measuring oral proficiency in distance, face-to-face, and blended classrooms,* Language Learning & Technology **12**, 114 (2008).

[65] D. R. Garrison and N. Vaughan, *Blended learning in higher education* (Jossey-Bass, San Fransisco, 2008).

[66] M. Driessen, J. van Emmerik, K. Fuhri, L. Nygren-Junkin, and M. Spotti, *ICT Use in L2 Education for Adult Migrants: A qualitative study in the Netherlands and Sweden*, Tech. Rep. (2011).

[67] P. Steehouder and M. Tijssen, *Opbrengsten in Beeld. Rapportage Aanvalsplan Laaggeletterdheid 2006-2010*, Tech. Rep. (CINOP, Den Bosch, 2011).

[68] W. Houtkoop, J. Allen, M. Buisman, D. Fouarge, and R. van der Velden, *Kernvaardigheden in Nederland: Resultaten van de Adult Literacy and Life Skills Survey (ALL)* (Expertisecentrum Beroepsonderwijs, 2012).

[69] P. J. Stappers, F. S. Visser, and I. Keller, *The Role of Prototypes and Frameworks for Structuring Explorations by Research Through Design,* in *The Routledge Companion to Design Research* (2014) pp. 163–174.

[70] S. D. Tripp and B. Bichelmeyer, *Rapid prototyping: an alternative instructional design strategy,* Educational Technology Research and Development **38**, 31 (1990).

[71] A. H. M. Cremers, Y. J. F. M. Jansen, M. A. Neerincx, D. G. M. Schouten, and A. Kayal, *Inclusive Design and Anthropological Methods to Create Technological Support for Societal Inclusion,* in *Universal Access in Human-Computer Interaction. Design and Development Methods for Universal Access*, edited by C. Stephanidis and M. Antona (2014) pp. 31–42.

[72] W. W. Gaver, T. Dunne, and E. Pacenti, *Cultural Probes,* Interactions **6**, 21 (1999).

[73] M. Westera, J. Boschloo, J. van Diggelen, L. S. Koelewijn, M. A. Neerincx, and N. J. J. M. Smets, *Employing use-cases for piecewise evaluation of requirements and claims,* in *Proceedings of the 28th Annual European Conference on Cognitive Ergonomics* (2010) p. 8.

[74] B. G. Glaser and A. Strauss, *The Discovery of Grounded Theory: Strategies for Qualitative Research* (Transaction Publishers, 1967).

[75] F. S. Visser, *Bringing the everyday life of people into design*, Ph.D. thesis (2009).

[76] K. Rodil, K. L. Jensen, M. Rehm, and H. Winschiers-Theophilus, *Identifying and Representing Elements of Local Contexts in Namibia,* in *Human-Computer Interaction, Part III, HCII 2013, LNCS 8006* (2013) pp. 332–341.

[77] M. Hanekamp, *Vijftien jaar werken met ambassadeurs,* Alfa-nieuws , 4 (2012).

[78] D. G. M. Schouten, N. J. J. M. Smets, M. Driessen, K. Fuhri, M. A. Neerincx, and A. H. M. Cremers, *Requirements for a virtual environment to support the social participation education of low-literates,* Universal Access in the Information Society **16**, 681 (2016).

[79] W. Bank, *Inclusion Matters: The Foundation for Shared Prosperity* (Washtington, DC, 2013).

[80] N. Hoeymans, J. M. Timmermans, M. M. Y. de Klerk, A. H. de Boer, D. J. H. Deeg, J. L. Poppelaars, F. Thissen, J. C. D. Fortuijn, and A. E. M. de Hollander, *Gezond actief: de relatie tussen ziekten, beperkingen en maatschappelijke participatie onder Nederlandse ouderen*, Tech. Rep. (RIVM, 2005).

[81] G. Pauwels and J. Pickery, *Wie participeert niet? Ongelijke deelname aan het maatschappelijke leven in verschillende domeinen,* SVR-rapport 2007/5 (2007).

[82] S. van Linden and A. H. M. Cremers, *Cognitive Abilities of Functionally Illiterate Persons Relevant to ICT Use,* in *ICCHP 2008, LNCS 5105* (2008) pp. 705–712.

[83] T. Monahan, G. McArdle, and M. Bertolotto, *Virtual reality for collaborative e-learning,* Computers & Education **50**, 1339 (2008).

[84] M. Neerincx, W. Vught, O. Blanson Henkemans, E. Oleari, J. Broekens, R. Peters, F. Kaptein, Y. Demiris, B. Kiefer, D. Fumagalli, *et al.*, *Socio-cognitive engineering of a robotic partner for child's diabetes self-management,* Frontiers in Robotics and AI **6**, 118 (2019).

[85] M. A. Neerincx and J. Lindenberg, *Situated cognitive engineering for complex task environments,* in *Naturalistic decision making and macrocognition*, edited by J. M. Schraagen, L. G. Militello, T. Ormerod, and R. Lipshitz (Ashgate Publishing Limited, Aldershot, UK, 2008) pp. 373–390.

[86] J. Janssen, J. G. M. de Jong, and R. T. Paulissen, *Mind Map Operational Demands & Paper Prototype of Operational Demands Module in sCET*, Tech. Rep. (TNO, Soesterberg, the Netherlands, 2015).

[87] A. Broos and K. Roe, *The digital divide in the playstation generation: Self-efficacy, locus of control and ICT adoption among adolescents,* Poetics **34**, 306 (2006).

[88] M. S. Eastin and R. LaRose, *Internet self-efficacy and the psychology of the digital divide,* Journal of Computer-Mediated Communication **6**, 0 (2000).

[89] J. Kurvers, K. Dalderop, and W. Stockmann, *Cursistprofielen Laaggeletterdheid NT1 & NT2*, Tech. Rep. (Steunpunt Taal en Rekenen VE, Tilburg, 2013).

[90] N. Verhelst, P. van Avermaet, S. Takala, N. Figueras, and B. North, *Common European Framework of Reference for Languages: learning, teaching, assessment* (Cambridge University Press, 2009).

[91] D. G. M. Schouten, *Improving Societal Participation for Low-Literates and Non-Natives Through Computer Support Tool Design,* (2013).

[92] M. P. Breen and C. N. Candlin, *The essentials of a communicative curriculum in language teaching,* Applied linguistics **1**, 89 (1980).

[93] J. S. Vacca, *Using Scaffolding Techniques to Teach a Social Studies Lesson About Buddha to Sixth Graders,* Journal of Adolescent & Adult Literacy **51**, 652 (2008).

[94] S. Cassidy, *Learning styles: An overview of theories, models, and measures,* Educational Psychology **24**, 419 (2004).

[95] T. Garcia and P. Pintrich, *Assessing students' motivation and learning strategies in the classroom context: The motivated strategies for learning questionnaire,* in *Alternatives in assessment of achievements, learning processes and prior knowledge* (Springer Netherlands, 1996) pp. 319–339.

[96] S. B. Merriam, *Andragogy and Self-Directed Learning: Pillars of Adult Learning Theory,* in *New Directions for Adult and Continuing Education 89* (Jossey-Bass, 2001) Chap. 1, pp. 3–13.

[97] M. M. M. Peeters, J.-J. C. Meyer, K. van den Bosch, and M. A. Neerincx, *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 (2012) pp. 266–272.

[98] A. Bandura, *Self-efficacy: toward a unifying theory of behavioral change.* Psychological review **84**, 191 (1977).

[99] A. Bandura, *Social foundations of thought and action* (Prentice Hall, Englewood Cliffs, New Jersey, 1986).

[100] D. H. Schunk, *Goal setting and self-efficacy during self-regulated learning,* Educational psychologist **25**, 71 (1990).

[101] D. H. Schunk, *Self-efficacy and academic motivation,* Educational psychologist **26**, 207 (1991).

[102] D. H. Schunk, *Self-efficacy, motivation, and performance,* Journal of Applied Sport Psychology **7**, 112 (1995).

[103] D. H. Schunk, *Self-Efficacy for Reading and Writing: Influence of Modeling, Goal Setting, and Self-Evaluation,* Reading and Writing Quarterly **19**, 159 (2003).

[104] P. R. Albion, *Self-Efficacy Beliefs as an Indicator of Teachers' Preparedness for Teaching with Technology,* in *Proceedings of the 10th International Conference of the Society for Information Technology & Teacher Education (SITE 1999)* (Association for the Advancement of Computing in Education (AACE), 1999) pp. 1602–1608.

[105] R. D. Johnson, S. Hornik, and E. Salas, *An empirical examination of factors contributing to the creation of successful e-learning environments,* International Journal of Human-Computer Studies **66**, 356 (2008).

[106] A. H. M. Cremers, J. G. M. de Jong, and J. S. van Balken, *User-centered design with illiterate persons: the case of the ATM user interface,* in *ICCHP 2008, LNCS 5105* (2008) pp. 713–720.

[107] D. G. M. Schouten, *Computer-supported learning of societal participation skills by low-literates and non-natives,* in *Proceedings of the 2014 European Conference on Cognitive Ergonomics* (ACM, 2014) p. 34.

[108] T. C. Clapper, *Beyond Knowles: What Those Conducting Simulation Need to Know About Adult Learning Theory,* Clinical Simulation in Nursing **6**, e7 (2010).

[109] A. Bandura, *Self-efficacy: The exercise of control* (W.H. Freeman, New York, 1997).

[110] C. Sansone, J. L. Smith, D. B. Thoman, and A. MacNamara, *Regulating interest when learning online: Potential motivation and performance trade-offs,* The Internet and Higher Education **15**, 141 (2012).

[111] M. B. Carvalho, *Designing for Low-literacy Users: a Framework for Analysis of User-Centred Design Methods*, Ph.D. thesis (2011).

[112] A. H. M. Cremers, J. G. M. de Jong, J. M. Kessens, S. van Linden, and R. Looije, *Automated Teller Machines for illiterates: target group and domain analysis*, Tech. Rep. (2008).

[113] F. Guay, R. J. Vallerand, and C. Blanchard, *On the Assessment of Situational Intrinsic and Extrinsic Motivation: The Situational Motivation Scale (SIMS),* Motivation and Emotion **24**, 175 (2000).

[114] J. Gorges and C. Kandler, *Adults' learning motivation: Expectancy of success, value, and the role of affective memories,* Learning and Individual Differences **22**, 610 (2012).

[115] E. A. Linnenbrink and P. R. Pintrich, *Motivation as an Enabler for Academic Success,* School Psychology Review **31**, 313 (2002).

[116] S. J. Warren, M. J. Dondlinger, and S. A. Barab, *A MUVE Towards PBL Writing: Effects of a Digital Learning Environment Designed To Improve Elementary Student Writing,* Journal of Research on Technology in Education **41**, 113 (2008).

[117] T. S. H. Teo, V. K. G. Lim, and R. Y. C. Lai, *Intrinsic and extrinsic motivation in Internet usage,* Omega: The International Journal of Management Science **27**, 25 (1999).

[118] K. Facer, R. Joiner, D. Stanton, J. Reid, R. Hull, and D. Kirk, *Savannah: mobile gaming and learning?* Journal of Computer Assisted Learning , 399 (2004).

[119] Y. Levy, *Comparing dropouts and persistence in e-learning courses,* Computers & Education **48**, 185 (2007).

[120] A. Parker, *Identifying predictors of academic persistence in distance education,* USDLA Journal **17**, 55 (2003).

[121] G. J. Conti, *Development of a user-friendly instrument for identifying the learning strategy preferences of adults,* Teaching and Teacher Education **25**, 887 (2009).

[122] S. B. Merriam, *Adult Learning Theory for the Twenty-First Century,* New Directions for Adult and Continuing Education 119 , 93 (2008).

[123] M. S. Knowles, *Andragogy, not pedagogy,* Adult leadership **16** (1968).

[124] M. S. Knowles, *What Is Andragogy?* in *The Modern Practice of Adult Education: From Pedagogy to Andragogy. Revised and updated.* (Prentice Hall Regents, Englewood Cliffs, New Jersey, 1970) Chap. 4, pp. 40–59.

[125] M. S. Knowles, *The Adult Learner: A Neglected Species*, 1 (Gulf Publishing Company, Houston, Texas, 1973).

[126] J. Mezirow, *Learning to Think Like an Adult: Core Concepts of Transformation Theory,* in *Learning as Transformation. Critical Perspectives on a Theory in Progress*, edited by S. F. Jossey-Bass (2000) Chap. 1, pp. 3–33.

[127] K. Illeris, *Characteristics of Adult Learning,* in *Adult Education - Adult Learning, Instruction and Program Planning*, 1 (2010) pp. 36–41.

[128] B. Kim, *Social Constructivism,* Emerging perspectives on learning, teaching, and technology **1**, 16 (2001).

[129] C. A. Hansman, *Context-Based Adult Learning,* in *New Directions for Adult and Continuing Education 89* (Jossey-Bass, 2001) Chap. 5, pp. 43–51.

[130] J. S. Brown, A. Collins, and P. Duguid, *Situated cognition and the culture of learning,* Educational researcher **18**, 32 (1989).

[131] J. Lave and E. Wenger, *Situated learning: Legitimate peripheral participation* (Cambridge university press, 1991).

[132] J. M. Dirkx, *The Power of Feelings: Emotion, Imagination, and the Construction of Meaning in Adult Learning,* in *New Directions for Adult and Continuing Education 89* (Jossey-Bass, 2001) Chap. 7, pp. 63–72.

[133] L. H. Hill, *The Brain and Consciousness: Sources of Information for Understanding Adult Learning,* in *New Directions for Adult and Continuing Education 89* (Jossey-Bass, 2001) Chap. 8, pp. 73–81.

[134] D. R. Garrison and W. Archer, *A transactional perspective on teaching and learning: A framework for adult and higher education* (Pergamon, Oxford, UK, 2000).

[135] D. H. Jonassen, *Thinking Technology: Toward a Constructivist Design Model,* Educational Technology **34**, 34 (1994).

[136] D. H. Jonassen, M. Davidson, M. Collins, J. Campbell, and B. B. Haag, *Constructivism and computer□mediated communication in distance education,* American Journal of Distance Education **9**, 7 (1995).

[137] L. Cuban, *Oversold and underused: Computers in the classroom* (Harvard University Press, 2001).

[138] A. J. A. M. van Deursen and J. A. G. M. van Dijk, *Trendrapport Computer- en Internetgebruik 2011. Een Nederlands en Europees perspectief*, Tech. Rep. (Universiteit Twente, Enschede, 2011).

[139] K. Kiili, *Digital game-based learning: Towards an experiential gaming model,* The Internet and Higher Education **8**, 13 (2005).

[140] C. Petranek, *A maturation in experiential learning: Principles of simulation and gaming,* Simulation & Gaming **25**, 513 (1994).

[141] D. C. Thatcher, *Promoting learning through games and simulations,* Simulation & Gaming **21**, 262 (1990).

[142] A. Doshi, *How gaming could improve information literacy,* Computers in libraries **26**, 14 (2006).

[143] F. Ke, *A case study of computer gaming for math: Engaged learning from gameplay?* Computers & Education **51**, 1609 (2008).

[144] L. Ermi and F. Mäyrä, *Fundamental Components of the Gameplay Experience: Analysing Immersion,* in *Proceedings of DiGRA 2005 Conference: Changing Views – Worlds in Play* (2005) p. 14.

[145] M. D. Dickey, *Game design and learning: a conjectural analysis of how massively multiple online role-playing games (MMORPGs) foster intrinsic motivation,* Educational Technology Research and Development **55**, 253 (2007).

[146] L. P. Rieber, *Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games,* Educational Technology Research and Development **44**, 43 (1996).

[147] W. C. Kriz, *Creating Effective Learning Environments and Learning Organizations through Gaming Simulation Design,* Simulation & Gaming **34**, 495 (2003).

[148] A. Astell, N. Alm, R. Dye, G. Gowans, P. Vaughan, and M. Ellis, *Digital Video Games for Older Adults with Cognitive Impairment,* in *ICCHP 2014, Part I, LNCS 8547*, edited by K. Miesenberger, D. Fels, D. Archambault, P. Penáz, and W. Zagler (Springer, Paris, 2014) pp. 264–271.

[149] S. J. Czaja, R. Schulz, D. Perdomo, and S. N. Nair, *The Feasability and Efficacy of Technology-Based Support Groups among Family Caregivers of Persons with Dementia,* in *ICCHP 2014, Part I, LNCS 8547*, edited by K. Miesenberger, D. Fels, D. Archambault, P. Penáz, and W. Zagler (Springer, Paris, 2014) pp. 455–458.

[150] D. G. M. Schouten, I. Pfab, A. H. M. Cremers, B. van Dijk, and M. A. Neerincx, *Gamification for low-literates: Findings on motivation, user experience, and study design,* in *ICCHP 2014, Part I, LNCS 8547*, edited by K. Miesenberger, D. Fels, D. Archambault, P. Penáz, and W. Zagler (Springer, Paris, 2014) pp. 494–501.

[151] C. Bouras, E. Giannaka, and T. Tsiatsos, *Virtual collaboration spaces: the EVE community,* in *Proceedings of the 2003 Symposium on Applications and the Internet (SAINT)* (2003) pp. 48–55.

[152] S. de Freitas and F. Liarokapis, *Serious Games: A New Paradigm for Education?* in *Serious games and edutainment applications*, edited by M. Ma, A. Oikonomou, and L. C. Jain (Springer London, London, 2011) pp. 9–23.

[153] J. Cassell and T. W. Bickmore, *Negotiated collusion: Modeling social language and its relationship effects in intelligent agents,* User Modeling and User-Adapted Interaction , 89 (2003).

[154] L. A. Annetta, *Video games in education: Why they should be used and how they are being used,* Theory Into Practice **47**, 229 (2008).

[155] K.-C. Chen and S.-J. Jang, *Motivation in online learning: Testing a model of self-determination theory,* Computers in Human Behavior **26**, 741 (2010).

[156] D. G. M. Schouten, R. T. Paulissen, M. Hanekamp, A. Groot, M. A. Neerincx, and A. H. M. Cremers, *Low-literates' support needs for societal participation learning: Empirical grounding of theory- and model-based design,* Cognitive Systems Research **45**, 30 (2017).

[157] K. Illeris, *Towards a contemporary and comprehensive theory of learning,* International Journal of Lifelong Education **22**, 396 (2003).

[158] K. M. Bass and R. Glaser, *CSE Report*, Tech. Rep. (2004).

[159] S.-S. Liaw, *Investigating students' perceived satisfaction, behavioral intention, and effectiveness of e-learning: A case study of the Blackboard system,* Computers and Education **51**, 864 (2008).

[160] D. Zhang, L. Zhou, R. O. Briggs, and J. F. Nunamaker, *Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness,* Information and Management **43**, 15 (2006).

[161] J. Biggs and P. Moore, *The process of learning*, 3rd ed. (Prentice Hall, Australia, 1993).

[162] P. Cybinski and S. Selvanathan, *Learning Experience and Learning Effectiveness in Undergraduate Statistics: Modeling Performance in Traditional and Flexible Learning Environments,* Decision Sciences Journal of Innovative Education **3**, 251 (2005).

[163] M. A. Neerincx, *Situated cognitive engineering for crew support in space,* Personal and Ubiquitous Computing **15**, 445 (2011).

[164] W. W. Gaver, A. Boucher, S. Pennington, and B. Walker, *Cultural probes and the value of uncertainty,* Interactions-Funology **11**, 53 (2004).

[165] M. Hanekamp, *Werken met cultural probes: Technologie als hulpmiddel bij het opsporen van voor deelnemers 'lastige' situates,* Les 187 , 12 (2014).

[166] S. Schmehl, S. Deutsch, J. Schrammel, L. Paletta, and M. Tscheligi, *Directed Cultural Probes: Detecting Barriers in the Usage of Public Transportation,* in *INTERACT 2011, Part I, LNCS 6946* (2011) pp. 404–411.

[167] K. Charmaz, *Grounded Theory,* in *Rethinking Methods in Psychology*, edited by J. A. Smith, R. Harré, and L. van Langenhove (1996) Chap. 3, pp. 27–49.

[168] B. G. Glaser, *Conceptualization: On Theory and Theorizing Using Grounded Theory.* International Journal of Qualitative Methods **1**, 23 (2002).

[169] A. Strauss and J. M. Corbin, *Grounded theory methodology: an overview,* in *Handbook of Qualitative Research* (1994) Chap. 17, pp. 273–285.

[170] K. Charmaz, *Grounded Theory as an Emergent Method,* in *Handbook of Emergent Methods*, 3, edited by S. N. Hesse-Biber and P. Leavy (New York: The Guilford Press, 2008) Chap. 7, pp. 155–172.

[171] J. M. Corbin and A. Strauss, *Grounded theory research: Procedures, canons, and evaluative criteria,* Zeitschrift für Soziologie **19**, 418 (1990).

[172] B. G. Glaser and J. Holton, *Remodeling Grounded Theory,* (2004).

[173] R. Suddaby, *From the editors: what Grounded Theory is not,* Academy of Management Journal **49**, 633 (2006).

[174] Y. S. Lincoln and E. G. Guba, *Naturalistic inquiry*, vol. 75 ed. (Sage, 1985).

[175] A. Strauss and J. M. Corbin, *Basics of qualitative research: Procedures and techniques for developing grounded theory*, 2nd ed. (Sage, Thousand Oaks, CA, 1998).

[176] H. Zoon, A. H. M. Cremers, and B. Eggen, *'Include', a Toolbox of User Research for Inclusive Design,* in *Chi Sparks 2014* (2014) pp. 113–121.

[177] I. F. Alexander and N. Maiden, *Scenarios, stories, use cases: through the systems development life-cycle* (John Wiley & Sons, 2005).

[178] J. Mezirow, *Transformative Learning: Theory to Practice,* in *New Directions for Adult and Continuing Education 74*, 74 (1997) pp. 5–12.

[179] D. Maulsby, S. Greenberg, and R. Mander, *Prototyping an intelligent agent through Wizard of Oz,* in *ACM SIGCHI Conference on Human Factors in Computing Systems* (Amsterdam, 1993) pp. 277–284.

[180] D. G. M. Schouten, N. J. J. M. Smets, M. Driessen, M. Hanekamp, A. H. M. Cremers, and M. A. Neerincx, *User Requirement Analysis of Social Conventions Learning Applications for Non-Natives and Low-Literates,* in *Engineering Psychology and Cognitive Ergonomics. Understanding Human Cognition* (Springer, 2013) pp. 354–363.

[181] L. L. Kramer, S. ter Stal, B. C. Mulder, E. de Vet, and L. van Velsen, *Developing embodied conversational agents for coaching people in a healthy lifestyle: Scoping review,* Journal of Medical Internet Research **22**, 1 (2020).

[182] S. ter Stal, L. L. Kramer, M. Tabak, H. op den Akker, and H. Hermens, *Design Features of Embodied Conversational Agents in eHealth: a Literature Review,* International Journal of Human-Computer Studies **138**, 102409 (2020).

[183] D. G. M. Schouten, F. Venneker, T. Bosse, M. A. Neerincx, and A. H. M. Cremers, *A Digital Coach That Provides Affective And Social Learning Support to Low-Literate Learners,* IEEE Transactions on Learning Technologies **11**, 67 (2017).

[184] B. Alsharbi and D. Richards, *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* (2017) pp. 11–15.

[185] H. C. van Vugt, E. A. Konijn, J. F. Hoorn, and J. Veldhuis, *Why fat interface characters are better e-health advisors,* in *International Workshop on Intelligent Virtual Agents* (Springer, 2006) pp. 1–13.

[186] R. Khan and A. De Angeli, *The attractiveness stereotype in the evaluation of embodied conversational agents,* in *IFIP Conference on Human-Computer Interaction* (Springer, 2009) pp. 85–97.

[187] A. de Angeli and S. Brahnam, *Sex Stereotypes and Conversational Agents,* AVI 2006 Gender and Interaction: Real and Virtual Women in a Male World Workshop paper (2006).

[188] S. Zhou, Z. Zhang, and T. Bickmore, *Adapting a persuasive conversational agent for the chinese culture,* in *2017 International Conference on Culture and Computing (Culture and Computing)* (IEEE, 2017) pp. 89–96.

[189] B. J. Fogg, *Persuasive Technology: Using computers to change what we think and do* (Morgan Kaufman, San Francisco, 2002).

[190] M. ter Maat, K. P. Truong, and D. Heylen, *How agents' turn-taking strategies influence impressions and response behaviours,* Presence: Teleoperators and Virtual Environments **20**, 412 (2011).

[191] T. W. Bickmore, D. Schulman, and C. L. Sidner, *Automated interventions for multiple health behaviors using conversational agents.* Patient education and counseling **92**, 142 (2013).

[192] F. de Rosis, N. Novielli, V. Carofiglio, A. Cavalluzzi, and B. De Carolis, *User modeling and adaptation in health promotion dialogs with an animated character,* Journal of biomedical informatics **39**, 514 (2006).

[193] A. Shamekhi, H. Trinh, T. W. Bickmore, T. R. DeAngelis, T. Ellis, B. V. Houlihan, and N. K. Latham, *A virtual self-care coach for individuals with spinal cord injury,* in *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility* (ACM, 2016) pp. 327–328.

[194] E. Hudlicka, *Virtual coach for mindfulness meditation training,* in *2011 AAAI Spring Symposium Series* (2011).

[195] G. Veletsianos and C. Miller, *Conversing with pedagogical agents: A phenomological exploration of interacting with digital entities,* British Journal of Educational Technology **39**, 969 (2008).

[196] A. K. Mahmood and E. Ferneley, *The use of animated agents in e□learning environments: an exploratory, interpretive case study,* ALT-J, Research in Learning Technology **14**, 153 (2006).

[197] Y. Miao, J. Engler, A. Giemza, S. Weinbrenner, and H. U. Hoppe, *Develoment of a process-oriented scaffolding agent in an open-ended inquiry learning environment,* Research and Practice in Technology Enhanced Learning **7**, 105 (2012).

[198] E. Shaw, R. Ganeshan, W. L. Johnson, and D. Millar, *Building a case for agent-assisted learning as a catalyst for curriculum reform in medical education,* Frontiers in Artificial Intelligence and Applications **50**, 509 (1999).

[199] J. C. Lester, S. A. Converse, S. E. Kahler, S. T. Barlow, B. A. Stone, and R. S. Bhogal, *The persona effect: affective impact of animated pedagogical agents,* in *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems* (1997).

[200] T. W. Bickmore and J. Cassell, *Social Dialogue with Embodied Conversational Agents,* Advances in Natural Multimodal Dialogue Systems **30**, 23 (2005).

[201] R. Moreno, R. E. Mayer, H. A. Spires, and J. C. Lester, *The Case for Social Agency in Computer-Based Teaching: Do Students Learn More Deeply When They Interact With Animated Pedagogical Agents,* Cognition and Instruction **19**, 177 (2001).

[202] T. W. Bickmore and R. W. Picard, *Establishing and Maintaining Long-Term Human-Computer Relationships,* ACM Transactions on Computer-Human Interaction **12**, 293 (2005).

[203] S.-H. Kang and J. Gratch, *People like virtual counselors that highly-disclose about themselves.* Annual Review of Cybertherapy and Telemedicine **167** (2011).

[204] J. van de Pol and E. Elbers, *Scaffolding student learning: A micro-analysis of teacher–student interaction,* Learning, Culture and Social Interaction **2**, 32 (2013).

[205] J. Echevarria, M. Vogt, and D. Short, *Making content comprehensible for English learners: The SIOP model* (Allyn and Bacon, Boston, 2004).

[206] T. Fortune, *Scaffolding Techniques in CBI Classrooms*, Tech. Rep. (2004).

[207] W. R. Miller and S. Rollnick, *Ten things that motivational interviewing is not,* Behavioural and cognitive psychotherapy **37**, 129 (2009).

[208] O. A. Blanson Henkemans, P. J. M. van der Boog, J. Lindenberg, C. A. P. G. van der Mast, M. A. Neerincx, and B. J. H. M. Zwetsloot-Schonk, *An online lifestyle diary with a persuasive computer assistant providing feedback on self-management,* Technology and Health Care **17**, 253 (2009).

[209] S. A. H. Friederichs, A. Oenema, C. Bolman, J. Guyaux, H. M. van Keulen, and L. Lechner, *I Move: systematic development of a web-based computer tailored physical activity intervention, based on motivational interviewing and self-determination theory,* BMC Public Health **14** (2014).

[210] R. Looije, M. A. Neerincx, and F. Cnossen, *Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors,* International Journal of Human-Computer Studies **68**, 386 (2010).

[211] L. Sobell and M. Sobell, *Motivational interviewing strategies and techniques: rationales and examples,* (2008).

[212] J. Cassell and T. W. Bickmore, *External Manifestations of Trustworthiness in the Interface,* Communications of the ACM **43**, 50 (2000).

[213] M. Bayles, *Online Banking: Why People Are Branching Out,* Transfer **35**, 90 (2004).

[214] J. Nielsen, *Ten Usability Heuristics,* Communications of the ACM **3**, 1 (2005).

[215] M. O. Leavitt and B. Schneiderman, *US Department of Health and Human Services*, Tech. Rep. (2006).

[216] M. De Jong, M. Theune, and D. Hofs, *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* (2008) pp. 207–214.

[217] A. A. Deneka, *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).

[218] S. ter Stal, M. Tabak, H. op den Akker, T. Beinema, and H. Hermens, *Who do you prefer? the effect of age, gender and role on users' first impressions of embodied conversational agents in ehealth,* International Journal of Human–Computer Interaction **36**, 881 (2020).

[219] B. S. Bloom, *Taxonomy of Educational Objectives, Handbook I: The Cognitive Domain* (Longman, 1956).

[220] M. Bercht and R. M. Viccardi, *Pedagogical agents with affective and cognitive dimensions,* in *Proceedings of 5th Ibero-American Congress on Educational Informatics (RIBIE 2000)* (Nina del Mar, Chile, 2000).

[221] M. Haake and A. Gulz, *Visual realism and virtual pedagogical agents,* in *Proceedings of the 3rd International Design and Engagability Conference at NordiChi 2006 (iDec3)*, edited by J. Knight (2006).

[222] D. G. M. Schouten, P. Massink, S. F. Donker, M. A. Neerincx, and A. H. M. Cremers, *User Modeling and User-Adapted Interaction* (Springer Netherlands, 2020) pp. 1–41.

[223] D. G. M. Schouten, A. A. Deneka, M. Theune, M. A. Neerincx, and A. H. M. Cremers, *An Embodied Conversational Agent Coach To Support Low-Literate Societal Participation Learning: Design, Development, and Evaluation,* Under review (2020).

[224] G. Fischer, *User modeling in human–computer interaction,* User modeling and user-adapted interaction **11**, 65 (2001).

[225] C. Stephanidis, *Adaptive techniques for universal access,* User modeling and user-adapted interaction **11**, 159 (2001).

[226] E. J. Horvitz, J. S. Breese, D. Heckerman, D. Hovel, and K. Rommelse, *The lumiere project: Bayesian user modeling for inferring the goals and needs of software users,* arXiv preprint arXiv:1301.7385 (2013).

[227] J. van de Pol, M. Volman, and J. Beishuizen, *Scaffolding in Teacher–Student Interaction: A Decade of Research,* Educational Psychology Review **22**, 271 (2010).

[228] L. S. Vygotsky, *Mind in society: the development of higher psychological processes*, edited by M. Cole, V. John-Steiner, S. Scribner, and E. Souberman (Harvard University Press, Cambridge, Massachusetts, 1980).

[229] D. Wood and H. Wood, *Vygotsky, tutoring and learning,* Oxford review of Education **22**, 5 (1996).

[230] G. M. Johnson, *Instructionism and Constructivism: Reconciling Two Very Good Ideas,* (2005).

[231] A. C. Graesser, S. D'Mello, and W. Cade, *Instruction based on tutoring,* in *Handbook of Research on Learning and Instruction* (2011) pp. 408–426.

[232] A. C. Graesser and N. K. Person, *Question asking during tutoring,* American Educational Research Journal **31**, 104 (1994).

[233] D. Wood, *Scaffolding, contingent tutoring, and computer-supported learning,* International Journal of Artificial Intelligence in Education **12**, 280 (2001).

[234] S. D'Mello and A. C. Graesser, *AutoTutor and affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back,* ACM Transactions on Interactive Intelligent Systems (TiiS) **2**, 23 (2012).

[235] G. Gibbs, C. Simpson, D. James, and S. Fleming, *Learning and Teaching in Higher Education,* (2004).

[236] M. R. Lepper and M. Woolverton, *The Wisdom of Practice: Lessons Learned from the Study of Highly Effective Tutors,* in *Improving Academic Achievement*, edited by J. Aronson (Academic Press, New York, New York, USA, 2002) Chap. 7, pp. 135–158.

[237] P. K. Bhowmick, S. Sarkar, and A. Basu, *Ontology based user modeling for personalized information access.* IJCSA **7**, 1 (2010).

[238] B. Ciloglugil and M. M. Inceoglu, *User modeling for adaptive e-learning systems,* in *International Conference on Computational Science and Its Applications* (Springer, 2012) pp. 550–561.

[239] M. A. Tadlaoui, S. Aammou, M. Khaldi, and R. N. Carvalho, *Learner modeling in adaptive educational systems: a comparative study,* International Journal of Modern Education and Computer Science **8**, 1 (2016).

[240] G. Kaya and A. Altun, *A learner model for learning object based personalized learning environments,* in *Research Conference on Metadata and Semantic Research* (Springer, 2011) pp. 349–355.

[241] F. E. Benmarrakchi, J. El Kafi, and A. Elhore, *User modeling approach for dyslexic students in virtual learning environments,* International Journal of Cloud Applications and Computing (IJCAC) **7**, 1 (2017).

[242] B. Lehman, M. Matthews, and N. Person, *What Are You Feeling ? Investigating Student Affective States During Expert Human Tutoring Sessions,* , 50 (2008).

[243] E. C. Huskisson, *Visual analogue scales,* in *Pain measurement and assessment* (1983) pp. 33–37.

[244] M. M. Bradley and P. J. Lang, *Measuring Emotion: the Self-Assessment Manikin and the Semantic Differential,* Journal of Behavior Therapy & Experimental Psychiatry **25**, 49 (1994).

[245] K. F. Nimon, *Statistical assumptions of substantive analyses across the general linear model: A mini-review,* Frontiers in Psychology **3**, 1 (2012).

[246] P. Massink, *An intelligent tutoring system for low-literates to encourage societal participation*, Ph.D. thesis, Utrecht University (2015).

[247] D. Grandjean, D. Sander, and K. R. Scherer, *Conscious emotional experience emerges as a function of multilevel, appraisal-driven response synchronization,* Consciousness and Cognition **17**, 484 (2008).

[248] C. Darwin, P. Ekman, and P. Prodger, *The expression of the emotions in man and animals* (Oxford University Press, USA, 1998).

[249] P. Ekman, *An argument for basic emotions,* Cognition & Emotion **6**, 169 (1992).

[250] R. Plutchik, *The emotions* (University Press of America, 1991).

[251] W. G. Parrott, *Emotions in social psychology: Essential readings* (Psychology Press, 2001).

[252] S. Schachter and J. E. Singer, *Cognitive, social and physiological determinants of emotional state,* Psychological review6 **69**, 379 (1962).

[253] W. James, *The Principles of Psychology, Vol. 1* (Dover Publications, 1950).

[254] P. Ekman and W. V. Friesen, *Measuring facial movement,* Environmental psychology and nonverbal behaviour **1**, 56 (1976).

[255] P. C. Ellsworth and K. R. Scherer, *Appraisal processes in emotion,* in *Handbook of Affective Sciences* (2003) Chap. 29, p. 572.

[256] M. A. Nicolau, H. Gunes, and M. Pantic, *Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space,* IEEE Transactions on Affective Computing **2**, 92 (2011).

[257] J. A. Russel, *A Circumplex Model of Affect,* Journal of Personality and Social Psychology **39**, 1161 (1980).

[258] A. M. Oliveira, M. P. Teixeira, I. B. Fonseca, and M. Oliveira, *Joint Model-Parameter Validation of Self-Estimates of Valence and Aurosal: Probing a Differential-Weighting Model of Affective Intensity,* Proceedings of Fechner Day **22**, 245 (2006).

[259] P. A. Lewis, H. D. Critchley, P. Rotshtein, and R. J. Dolan, *Neural correlates of processing valence and arousal in affective words,* Cerebral cortex **17**, 742 (2007).

[260] J. A. Russel and A. Mehrabian, *Evidence for a three-factor theory of emotions,* Journal of research in Personality **11**, 273 (1977).

[261] W. R. Miller, *Motivational interviewing with problem drinkers,* Behavioural Psychotherapy **11**, 147 (1983).

[262] S. Rollnick and W. R. Miller, *What is motivational interviewing?* Behavioural and cognitive psychotherapy **23**, 325 (1995).

[263] K. M. Emmons and S. Rollnick, *Motivational interviewing in health care settings: opportunities and limitations,* American journal of preventive medicine **20**, 68 (2001).

[264] M. Feidakis, T. Daradoumis, and S. Caballé, *Emotion measurement in intelligent tutoring systems: what, when and how to measure,* in *Intelligent Networking and Collaborative Systems (INCoS), 2011 Third International Conference on* (IEEE, 2011) pp. 807–812.

[265] M. Wong, *Emotion assessment in evaluation of affective interfaces*, Master thesis, University of Waterloo, Ontario, Canada (2006).

[266] R. Pekrun, T. Goetz, A. C. Frenzel, P. Barchfeld, and R. P. Perry, *Measuring emotions in students' learning and performance: the Achievement Emotions Questionnaire (AEQ),* Contemporary educational psychology **36**, 36 (2011).

[267] P. G. Zimmermann, S. Guttormsen, B. Danuser, and P. Gomez, *Affective computing - a rationale for measuring mood with mouse and keyboard,* International journal of occupational safety and ergonomics **9**, 539 (2003).

[268] P. G. Zimmermann, *Beyond usability-measuring aspects of user experience*, Doctoral dissertation, Swiss Federal Institute of Technology Zurich (2008).

[269] A. Burns, B. R. Greene, M. J. McGrath, T. J. O'Shea, B. Kuris, S. M. Ayer, F. Stroiescu, and V. Cionca, *SHIMMER™ - A wireless sensor platform for noninvasive biomedical research,* IEEE Sensors Journal **10**, 1527 (2010).

[270] P. Lewinski, T. M. den Uyl, and C. Butler, *Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader,* Journal of Neuroscience, Psychology, and Economics **7**, 227 (2014).

[271] M. Courgeon, J.-C. Martin, and C. Jacquemin, *MARC: a Multimodal Affective and Reactive Character,* Proceedings of the Workshop on Affective Interaction on Natural Environment (2008).

[272] M. Haake and A. Gulz, *Visual Stereotypes and Virtual Pedagogical Agents,* Educational Technology and Society **11**, 1 (2008).

[273] N. Fourati, A. Richard, S. Caillou, N. Sabouret, J.-C. Martin, E. Chanoni, and C. Clavel, *Facial Expressions of Appraisals Displayed by a Virtual Storyteller for Children,* in *International Conference on Intelligent Virtual Agents* (2016) pp. 234–244.

[274] E. W. Taylor, *Fostering Mezirow's Transformative Learning Theory in the Adult Education Classroom: A Critical Review,* CJSAE/RCÉÉA **14**, 1 (2000).

[275] T. M. Duffy and D. H. Jonassen, *Constructivism and the Technology of Instruction: A Conversation* (Psychology Press, 1992).

[276] W.-P. Brinkman, D. Hartanto, N. Kang, D. de Vliegher, I. L. Kampmann, N. Morina, P. M. Emmelkamp, and M. A. Neerincx, *A virtual reality dialogue system for the treatment of social phobia,* in *CHI'12 Extended Abstracts on Human Factors in Computing Systems* (ACM, 2012) pp. 1099–1102.

[277] W. Houtkoop, *Basisvaardigheden in Nederland*, Tech. Rep. (Max Goote Kenniscentrum voor Beroepsonderwijs en Volwasseneneducatie, 1999).

[278] A. Jokiaho, B. May, M. Specht, and S. Stoyanov, *Barriers to using e learning in an advanced way,* International Journal of Advanced Corporate Learning (iJAC) **11**, 17 (2018).

[279] A. J. A. M. van Deursen and J. A. G. M. van Dijk, *Measuring digital skills. Performance tests of operational, formal, information and strategic Internet skills among the Dutch population.* in *58th Conference of the International Communication Association* (Montreal, Canada, 2008) pp. 1–25.

[280] C. Lisetti, U. Yasavur, C. de Leon, R. Amini, N. Rishe, and U. Visser, *Building an On-Demand Avatar-Based Health Intervention for Behavior Change,* in *Proceedings of the Twenty-Fifth International Florida Intelligence Research Society Conference*, Mi (2012) pp. 175–180.

[281] C. Lisetti, U. Yasavur, U. Visser, and N. Rishe, *Toward Conducting Motivational Interviewing with an On-Demand Clinician Avatar for Tailored Health Behavior Change Interventions,* Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare (2011), 10.4108/icst.pervasivehealth.2011.246078.

[282] N. Novielli, F. de Rosis, and I. Mazzotta, *User attitude towards an embodied conversational agent: Effects of the interaction mode,* Journal of Pragmatics **42**, 2385 (2010).

[283] N. Novielli, I. Mazzotta, B. De Carolis, and S. Pizzutilo, *Analysing user's reactions in advice-giving dialogues with a socially intelligent ECA.* Cognitive processing **13 Suppl 2**, 487 (2012).

[284] O. S. Iversen and C. Nielsen, *Using Digital Cultural Probes in Design with Children,* (2003).

[285] A. Bryant, A. Strauss, and R. Rorty, *Grounded Theory and Pragmatism: The Curious Case of Anselm Strauss,* Forum: Qualitative Social Research **10**, 1 (2009).

[286] M. Driessen, *Van blinde vlek naar heldere focus: E-coaching bij online leren,* Vakblad Profiel , 30 (2016).

[287] A. G. Gift, *Visual Analogue Scales: Measurement of Subjective Phenomena,* Nursing Research **38**, 286 (1989).

[288] H. van Laerhoven, H. J. van der Zaag-Loonen, and B. H. F. Derkx, *A comparison of Likert scale and visual analogue scales as response options in children's questionnaires,* Acta Paediatrica **93**, 830 (2004).

[289] C. Munteanu, H. Molyneaux, J. Maitland, D. McDonald, R. Leung, H. Fournier, and J. Lumsden, *Hidden in plain sight: low-literacy adults in a developed country overcoming social and educational challenges through mobile learning support tools,* Personal and Ubiquitous Computing **18**, 1455 (2013).

[290] R. van der Meer, P. Doelen, N. Kistemaker, and S. Boonen, *Herkennen, bereiken en activieren van de moeilijkere bereikbare doelgroep*, Tech. Rep. (Stichting Lezen & Schrijven, Lost Lemon, Muzus, 2019).

[291] Z. Wang, T. Schultz, and A. Waibel, *Comparison of acoustic model adaptation techniques on non-native speech,* in *International Conference on Acoustics, Speech, and Signal Processing 2003 (ICASSP '03)* (2003).

[292] J. Kovač, P. Peer, and F. Solina, *Human Skin Colour Clustering for Face Detection*, Vol. 2 (IEEE, 2003) pp. 144–148.

[293] V. Vezhnevets, V. Sazonov, and A. Andreeva, *A Survey on Pixel-Based Skin Color Detection Techniques,* Proc. Graphicon **3**, 85 (2003).

[294] I. Pavlidis and P. Symosek, *The Imaging Issue in an Automatic Face/Disguise Detection System,* in *Computer Vision Beyond the Visible Spectrum: Methods and Applications, 2000. Proceedings. IEEE Workshop on* (IEEE, 2000) pp. 15–24.

[295] A. Antonaci, R. Klemke, C. M. Stracke, and M. Specht, *Gamification in moocs to enhance users' goal achievement,* in *2017 IEEE Global Engineering Education Conference (EDUCON)* (IEEE, 2017) pp. 1654–1662.

[296] J. W. Brehm, *Post-decision changes in desirability of alternatives,* Journal of Abnormal and Social Psychology **52**, 384 (1956).

[297] M. K. Chen and J. L. Risen, *How Choice Affects and Reflects Preferences: Revisiting the Free-Choice Paradigm,* Journal of Personality and Social Psychology **99**, 573 (2010).

[298] A. Kayal, W.-P. Brinkman, M. A. Neerincx, and M. B. van Riemsdijk, *ePartners for socio-geographical support A normative framework approach,* (2012).

[299] P. M. Boynton and T. Greenhalgh, *Selecting, designing, and developing your questionnaire,* Bmj **328**, 1312 (2004).

[300] J. C. Jackson, J. Watts, T. R. Henry, J.-M. List, R. Forkel, P. J. Mucha, S. J. Greenhill, R. D. Gray, and K. A. Lindquist, *Emotion semantics show both cultural variation and universal structure.* Science (New York, N.Y.) **366**, 1517 (2019).

[301] M. M. M. Peeters, K. van den Bosch, J.-J. C. Meyer, and M. A. Neerincx, *The design and effect of automated directions during scenario-based training,* Computers & Education **70**, 173 (2014).

A

Software Overview

Presented below is detailed information on the software packages used for assessment in section 2.5.2. Included for each entry are the following data: name, functional description, year of production, and year of cancellation (if applicable). Each entry also indicates which of the three societal participation learning aspects the method is aimed at: language, skills, and/or knowledge of Dutch society.

EHBN

EHBN (Eerste Hulp Bij Nederland, meaning '*First Aid With The Netherlands*') is an older integration package that has been around since the 1990's. EHBN targets second language learners aiming for Dutch integration. Learning material consists of audio-supported multiple-choice questions. Language options are included.
Publisher: Malmberg
Year of production: 1990
Year of cancellation: 2013
Learning areas: Language, Knowledge of Dutch society

ETV.nl / Oefenen.nl

ETV.nl and Oefenen.nl ('*Practice.nl*') are two complementary websites that offer a large selection of learning programs. Among the programs offered are language learning segments aimed at native language learners, and integration courses aimed at second language learners. Multiple-choice questions are supported by video material.
Publisher: Expertise Foundation ETV.nl
Year of production: 2003
Year of cancellation: [*still in use*]
Learning areas: Language, Skills

Naar Nederland

Naar Nederland ('*To The Netherlands*') is described as the 'official self-study guide for the Dutch integration exam'. It targets second language learners. The method uses a DVD, several books, and online practice software. The complete package, including spoken and written segments, is offered in 18 languages. Speech recognition is also offered.
Publisher: Dutch Ministry of Social Affairs
Year of production: 2006
Year of cancellation: [*still in use*]
Learning areas: Language, Knowledge of Dutch society

Thuis in Nederlands

Thuis in Nederlands ('*At home in Dutch*') is a multimodal teaching method aimed at second language learners. It focuses on three core domains: Upbringing, Healthcare, and Education. It uses a mix of classroom book learning, practical assignments, roleplay, and e-learning; the latter component includes at-home work packages, a virtual 'participation board game', and a VR environment called the Virtual Neighbourhood.
Publisher: ITpreneurs
Year of production: 2008
Year of cancellation: 2014
Learning areas: Language, Skills, Knowledge of Dutch society

IJsbreker Plus

IJsbreker Plus ('*Icebreaker Plus*') is a language learning software package for second language learners looking for integration aid. The package combines independent online work with book exercises and classroom teaching. According to the website, the program offers a 'strong mix of learning types'. Audio-supported multiple choice questions are used, and different-language audio support is built in.
Publisher: ThiemeMeulenhoff
Year of production: 2010
Year of cancellation: [*still in use*]
Learning areas: Language, Knowledge of Dutch society

NL247

NL247 ('*NL Twentyfour-seven*') was developed by the same publisher as Thuis in Nederlands, and serves as a *de facto* sequel. NL247 supports low-literate second language learners in a wide variety of topics derived from the latest official Dutch integration exam. Different sets of exercises focus on reading, writing, comprehension, and vocabulary; learners are encouraged to focus on those skills areas they need most. NL247's software component is complemented by classroom lessons, written materials, and practical assignments.
Publisher: ITpreneurs
Year of production: 2014
Year of cancellation: [*still in use*]
Learning areas: Language, Skills, Knowledge of Dutch society

B

First Prototype Requirements Refinement

This appendix presents a description of all new requirements shown in Table 4.1, 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 ([228]; cf. [93, 301]). 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. [156]) 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 [89, 156]. 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 section 4.2.2.

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.

C

First Prototype Use Cases

Two use cases are provided here, based on the Online Banking and Service Desk exercises (section 4.2.3). 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* lists 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 Table 4.1's requirements, 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 section 4.1. 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. And '*conversation partner*' refers to the 'city hall service desk employee' character used in the service desk scenario.

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.

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.*"

<u>Post-conditions:</u>

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.

D

VESSEL Requirements Baseline (Formalized Cognitive Learning Support)

General Requirements	Coach Requirements	Exercise Requirements
R1. Adaptability. VESSEL should offer and/or support different learning 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.	**R1.1-C.** The coach should adapt its interaction style to individual user 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. *Cognitive support should be offered following the support rules model of prompt, explanation, hint, instruction, and modeling. And support should be offered at a learner-appropriate delay.*	**R1.1-E.** 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. *Exercises should always be challenging, and built on the assumption of coach support.*
R2. Sensitivity. 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.	**R2.1-C.** The coach should always address learners calmly and kindly, and avoid using phrases and broaching topics that upset low-literate learners.	**R2.1-E.** The exercises should be as sensitive as needed to reach the intended learning goals and difficulty level.
R3. Situatedness. 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. **R4. Collaboration.** 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.		**R3.1-E.** 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.
R5. Multimodality. 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.	**R5.1-C.** The coach should combine audio 'speech' with visual and textual supporting material.	**R5.1-E.** The exercises should be as multimodal as needed to reach the intended learning goals and difficulty level.
R6. Support. 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.	**R6.1-C. R6.1-C.** The coach should use *dialogue rules based on* verbal scaffolding to offer cognitive learning support. **R6.2-C.** The coach should use motivational interviewing techniques to offer affective learning support. **R6.3-C.** The coach should use small talk to offer social learning support.	
R7. Interactivity. 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.	**R7.1-C.** The coach should interact with users proactively by starting conversations and offering help, *following cognitive support rules, and according to a pre-defined timing scheme.* **R7-2-C.** The coach should interact with learners reactively by answering questions and demands for help. *The coach should only recognize and react to a particular set of pre-defined participant utterances, based on keywords.*	**R7.1-E.** The exercises should be interactive, requiring learners to use input mechanics to engage with the virtual environment in order to complete them.
R8. Gaming principles. 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.	

Table D.1: Refined VESSEL requirements baseline based on contingency rules. Unformatted text is the original description, text in *italics* has been added.

E

Recall Test Questions

1. On this picture of the internet banking website, what button should you press to reach the page where you can transfer money?
2. On this picture of a money transfer form, which important data has not been entered correctly?
3. On this picture of the internet banking website, what button should you press to reach the page where you can fill out a change of address?
4. On this picture of an address change form, which important data has not been entered correctly?
5. On this picture of the internet banking website, what button should you press to reach the page where you can move money to your savings account?
6. On this picture of a savings transfer form, which important data has not been entered correctly?

F

VESSEL Requirements Baseline (Formalized Cognitive, Affective, & Social Support)

General Requirements	Coach Requirements	Exercise Requirements
R1. Adaptability. 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.	**R1.1-C.** The coach should adapt its interaction style to individual user 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. Cognitive support should be offered following the support rules model of prompt, explanation, hint, instruction, and modeling. *Affective support should be offered based on the learner's affective state.* All support should be offered at a learner-appropriate delay.	**R1.1-E.** 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. Exercises should always be *cognitively and affectively* challenging, and built on the assumption of coach support.
R2. Sensitivity. 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.	**R2.1-C.** The coach should always address learners calmly and kindly, and avoid using phrases and broaching topics that upset low-literate learners.	**R2.1-E.** The exercises should be as sensitive as needed to reach the intended learning goals and difficulty level. *For exercises that include conversation partners, the conversation partner's dialogue should be as (in)sensitive as needed to reach the intended affective difficulty level.*
R3. Situatedness. 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.		**R3.1-E.** 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.
R4. Collaboration. 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.		
R5. Multimodality. 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.	**R5.1-C.** The coach should combine audio 'speech' with visual and textual supporting material.	**R5.1-E.** The exercises should be as multimodal as needed to reach the intended learning goals and difficulty level.
R6. Support. 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.	**R6.1-C.** The coach should use dialogue rules based on verbal scaffolding to offer cognitive learning support. **R6.2-C.** The coach should use *dialogue rules based on* motivational interviewing to offer affective learning support. **R6.3-C.** The coach should use *dialogue rules based on* small talk to offer social learning support.	
R7. Interactivity. 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.	**R7.1-C.** The coach should interact with users proactively by starting conversations and offering help, following cognitive, *affective, and social* support rules, *taking sensor data into account,* and according to a pre-defined timing scheme. **R7-2-C.** The coach should interact with learners reactively by answering questions and demands for help. The coach should only recognize and react to a particular set of pre-defined participant utterances, based on keywords.	**R7.1-E.** The exercises should be interactive, requiring learners to use input mechanics to engage with the virtual environment in order to complete them. *For exercises that include conversation partners, the conversation partner should deliver any cognitive support utterances.*
R8. Gaming principles. 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.	

Table F.1: Refined VESSEL requirements baseline. Requirements in table are based on formal cognitive/affective/social learning support rules. Unformatted text is the original description, text in *italics* has been added.

G

Small Talk Utterances

The utterances in Table G.1 (opposite page) are used as follows. In the first exercise, the coach uses utterances 1, 2, and 3. If the participant only responds utterance with a status (good/bad), utterance 4 or 5 is used. If the participant responds also immediately asks 'how are you doing', utterance 6 is used afterwards, otherwise it is skipped. Utterances 7 and 8a are used; if the participant indicates they do not know what 'volunteer work' means, a cognitive support utterance is used to explain it. Utterances 9 through 17 are used. The participant does the form part of the exercise. After this, utterances 18 and 19 are used. The participant then does the recruiter part of the exercise. Finally, utterance 20 is used. In the second exercise, the same order of utterances is used, with one difference: Utterance 8b is used instead of 8a (as we assume at this point the learner must know what 'volunteer work' means).

Order	Utterance (Dutch)	Utterance (English)
1	Hoi! Mijn naam is Anna. Ik help je bij deze oefening.	Hi! My name is Anna. I will help you with this exercise.
2	Hoe heet jij?	What is your name?
3	Leuk om je te ontmoeten. Hoe gaat het?	Nice to meet you. How are you doing?
4	Goed om te horen.	Good to hear.
5	Dat is jammer.	That's too bad.
6	Met mij gaat het goed. Dank je!	I'm doing fine. Thank you!
7	We gaan oefeningen doen in het leerprogramma. Ik ben er om je te helpen.	We're going to do learning exercises in the computer program. I am here to help you.
8a	In deze oefening ga je vrijwilligerswerk doen. Weet je wat dat is?	In this exercise, you will do volunteer work. Do you know what that is?
8b	In deze oefening ga je vrijwilligerswerk doen.	In this exercise, you will do volunteer work.
9	Heb je wel eens vrijwilligerswerk gedaan?	Have you ever done volunteer work before?
10	Vond je dat leuk?	Did you enjoy that?
11	Ik heb wel eens taalles gegeven aan kinderen. Dat was heel erg leuk.	I once taught language classes to children. That was very fun to do.
12	Ik vind dat vrijwilligerswerk heel belangrijk is. Wat vind jij?	I think volunteer work is very important. What do you think?
13	We gaan nu beginnen met de oefeningen.	We will now start the exercises.
14	We doen vandaag twee oefeningen. Eerst vul je een formulier in. En daarna ga je praten met mevrouw Peeters.	We will do two exercises today. First, you will fill out a form. Then, you will talk to miss Peeters.
15	Begrijp je wat ik zeg?	Do you understand what I'm saying?
16	Oke, laten we beginnen.	Alright, let's get started!
17	In deze oefening vul je een formulier in. Lees het formulier en vul het in.	In this exercise, you will fill out a form. Read the form, and fill it out.
18	We gaan nu naar de volgende oefening.	We will now move to the next exercise.
19	In deze oefening praat je met mevrouw Peeters. Deze mevrouw gaat jou vragen stellen over het formulier.	In this exercise, you will talk to miss Peeters. She will ask you questions about the form.
20	We zijn nu klaar met de oefeningen.	We are now done with the exercises.

Table G.1: Social support utterances used by ECA coach, shown in original Dutch (middle column) and translated English (right column).

H

Analysis Results Overview

The preceding tables (H.1, H.2, H.3) shows coach type evaluation results, order effects evaluation results, and post-hoc analysis results for 'learner type' and 'learner sex'. Per variable, the following is shown: Mean (standard deviation) for the two (ERQ/DM) or three (PAQ) measurement moments, the hypothesis that the variable was intended to measure. Analysis results (F-value, significance p, observer power β) for the Coach Type, Exercise Order, Learner Type, and Learner Sex, shown only if p<.05. And any significant interaction effects (further described in text). Grey boxes mean no value was measured or no test was conducted, blank boxes mean no significant result was found.

Var.	Hypo-thesis	Pre-Ex	Ex.1	Ex.2	Coach Type	Exercise Order	Learner Type	Learner Sex	Inter-action
Participant Assessment Questionnaire (PAQ)									
PAQ.1	H5	70.24 (*21.12*)	77.06 (*19.03*)	77.97 (*17.69*)		F=3.84 p=.03 β=.56			
PAQ.2	H5	58.53 (*25.04*)	64.79 (*21.78*)	66.94 (*22.81*)		F=5.64 p=.01 β=.83	F=4.17 p=.03 β=.69		
PAQ.3	H5	69.09 (*18.75*)	74.35 (*19.00*)	73.26 (*18.12*)					
PAQ.4	H5	71.21 (*25.54*)	71.91 (*24.11*)	70.41 (*24.09*)					*Ex. Ord.* **Lrn. Sex:* F=3.59 p=.04 β=.62
PAQ.5	H3	83.79 (*15.34*)	84.62 (*14.32*)	86.09 (*12.96*)					
PAQ.6	H2	36.59 (*36.58*)	35.56 (*34.97*)	36.82 (*35.63*)					
PAQ.7	H2	7.38 (*1.84*)	7.32 (*1.73*)	7.18 (*1.82*)			F=5.92 p=.02 β=.66		
PAQ.8	H2	5.24 (*2.52*)	5.41 (*2.41*)	4.79 (*2.54*)		F=4.75 p=.04 β=.56			
PAQ.9	H2	6.76 (*1.99*)	6.65 (*2.30*)	6.71 (*2.125*)					

Table H.1: Analysis results for the PAQ.

Var.	Hypo-thesis	Pre Ex	Ex.1	Ex.2	Coach Type	Exercise Order	Learner Type	Learner Sex	Inter-action
Exercise Reflection Questionnaire (ERQ)									
ERQ.1	H1		60.21 (*21.69*)	65.97 (*23.12*)			F=4.59 p=.04 β=.55		
ERQ.2	H1		59.65 (*22.55*)	59.53 (*22.10*)					*Ex. Ord. *Lrn. Type:* F=9.89 p=.00 β=.86
ERQ.3	H3		64.41 (*30.39*)	70.41 (*27.43*)					
ERQ.4	H6		.53 (*17.15*)	.65 (*20.82*)					
ERQ.5	H6		81.91 (*14.10*)	82.21 (*12.92*)					
ERQ.6	H6		74.12 (*20.09*)	76.21 (*22.86*)			F=7.51 p=.01 β=.76		

Table H.2: Analysis results for the ERQ.

Var.	Hypo-thesis	Pre Ex	Ex.1	Ex.2	Coach Type	Exercise Order	Learner Type	Learner Sex	Inter-action
Direct Measurements (DM)									
DM. 1-F	H1		384.8 (*206.0*)	289.4 (*149.3*)		F=16.04 p=.00 β=.97			
DM. 1-C	H1		291.1 (*84.99*)	271.9 (*94.84*)			F=10.39 p=.00 β=.87		
DM. 2-F	H1		5.00 (*5.86*)	2.38 (*3.76*)		F=8.40 p=.01 β=.80	F=4.59 p=.04 β=.55		
DM. 2-C	H1		1.97 (*2.68*)	1.59 (*2.56*)			F=5.35 p=.03 β=.61		
DM. 3-F	H1		1.06 (*2.72*)	1.19 (*2.53*)	F=14.43 p=.00 β=.96			F=9.33 p=.01 β=.84	
DM. 3-C	H1		2.25 (*3.20*)	3.03 (*4.55*)	F=52.76 p=.00 β=1.00				
DM. 4-F	H1		6.06 (*6.02*)	3.56 (*5.33*)	F=29.01 p=.00 β=1.00	F=5.05 p=.03 β=.59			*C. Type* **Lrn. Sex:* F=6.05 p=.02 β=.66
DM. 4-C	H1		4.22 (*3.67*)	4.63 (*5.97*)			F=5.46 p=.03 β=.62		
DM. 5-F	H6		1.56 (*2.09*)	1.69 (*2.98*)	F=8.48 p=.01 β=.81				
DM. 5-C	H6		2.84 (*3.45*)	1.56 (*2.49*)		F=9.78 p=.00 β=.86	F=8.59 p=.00 β=.85		
DM. 6	H4		1.99 (*1.00*)	2.31 (*1.35*)					

Table H.3: Analysis results for the DMs.

I

Questionnaire And Direct Variables

Variable	Description
Subjective measures: Participant Assessment Questionnaire (PAQ)	
PAQ.1 Self-efficacy (reading Dutch)	"I can read Dutch."
PAQ.2 Self-efficacy (computer use)	"I can use a computer."
PAQ.3 Self-efficacy (filling out a form)	"I can fill out a form."
PAQ.4 Self-efficacy (holding a conversation)	"I can hold a conversation with other people."
PAQ.5 Motivation (learning new things)	"I like to learn new things."
PAQ.6 Affect (fear of going to class)	"I think it's scary to go to class."
PAQ.7 Affect (valence)	"How good do you feel right now?"
PAQ.8 Affect (arousal)	"How active do you feel right now?"
PAQ.9 Affect (dominance)	"How strong do you feel right now?"
Subjective measures: Exercise Reflection Questionnaire (ERQ)	
ERQ.1 Performance (self-reported)	"I did the exercises well."
ERQ.2 Subjective mental effort	"It took effort to do the exercises well."
ERQ.3 Motivation (practice more)	"I would like to do these exercises more often."
ERQ.4 Coach opinion (performance)	"The coach helped me to do the exercises well."
ERQ.5 Coach opinion (trust)	"I trust, that the coach helps me well."
ERQ.6 Coach opinion (desire to retain)	"I would like to get help from this coach more often."
Objective measures: Direct measurements (*DM.x-F = form part of exercise, DM.x-C = conversation part of exercise*)	
DM.1-F Completion time (seconds)	Time from start of form part of exercise to completion.
DM.1-C Completion time (seconds)	Time from start of conversation part of exercise to completion.
DM.2-F Support utterances (cognitive)	Cognitive support utterances received in form part of exercise.
DM.2-C Support utterances (cognitive)	Cognitive support utterances received in conversation part of exercise.
DM.3-F Support utterances (affective)	Affective support utterances received in form part of exercise.
DM.3-C Support utterances (affective)	Affective support utterances received in conversation part of exercise.
DM.4-C Support utterances (total)	Total support utterances received in form part of exercise.
DM.4-F Support utterances (total)	Total support utterances received in conversation part of exercise.
DM.5-F Support utterances (self-initiated)	Learner-initiated coach interactions in form part of exercise.
DM.5-C Support utterances (self-initiated)	Learner-initiated coach interactions in conversation part of exercise.
DM.6 Recall test score	Score on end-of-experiment recall test.

Table I.1: Overview of quantitative measures. Includes: measure source (subjective measures from one of two questionnaires, or direct measurement) and description.

Curriculum Vitæ

Dylan Gerard Michel SCHOUTEN

03-08-1968	Born in Tilburg, the Netherlands.

Education

2004–2007	BSc Technical Innovation Sciences Eindhoven University of Technology, Eindhoven
2007–2011	MSc Human-Technology Interaction Eindhoven University of Technology, Eindhoven

Work Experience

2011–2012	Interface Designer at Euphoria-IT
2012–2016	PhD Researcher at Delft University of Technology & TNO
2016–2017	Postdoc Researcher at Twente University & T-Xchange
2017–2019	Postdoc Researcher at Northeastern University
2020	Associate Systems Designer at DIGIT Game Studios

List of Publications

8. **Schouten, D. G. M.**, Massink, P., Donker, S. F., Neerincx, M. A., Cremers, A. H. M. (2020). *Using scaffolding to formalize digital coach support for low-literate learners*. User Modeling and User-Adapted Interaction, 1-41.

7. **Schouten, D. G. M.**, Venneker, F., Bosse, T., Neerincx, M. A., & Cremers, A. H. M. (2017). A Digital Coach That Provides Affective And Social Learning Support to Low-Literate Learners. IEEE Transactions on Learning Technologies, **11 (1)**, 67–80.

6. **Schouten, D. G. M.**, Paulissen, R. T., Hanekamp, M., Groot, A., Neerincx, M. A., & Cremers, A. H. M. (2017). *Low-literates' support needs for societal participation learning: Empirical grounding of theory- and model-based design*. Cognitive Systems Research, **45**, 30–47.

5. **Schouten, D. G. M.**, Smets, N. J. J. M., Driessen, M., Fuhri, K., Neerincx, M. A., & Cremers, A. H. M. (2016). *Requirements for a virtual environment to support the social participation education of low-literates*. Universal Access in the Information Society, **16 (3)**, 681–698.

4. Cremers, A. H. M., Jansen, Y. J. F. M., Neerincx, M. A., **Schouten, D.G.M.**, & Kayal, A. (2014). *Inclusive design and anthropological methods to create technological support for societal inclusion*. In C. Stephanidis & M. Antona (Eds.), Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 8513 LNCS (Issue PART 1, pp. 31–42).

3. **Schouten, D. G. M.** (2014). *Computer-Supported Learning of Societal Participation Skills by Low-Literates and Non-Natives*. In Proceedings of the 2014 European Conference on Cognitive Ergonomics, 1–2.

2. **Schouten, D. G. M.**, Pfab, I., Cremers, A. H. M., van Dijk, B., & Neerincx, M. A. (2014). *Gamification for low-literates: Findings on motivation, user experience, and study design*. In K. Miesenberger, D. Fels, D. Archambault, P. Penáz, & W. Zagler (Eds.), ICCHP 2014, Part I, LNCS 8547 (pp. 494–501). Springer.

1. **Schouten, D. G. M.**, Smets, N., Driessen, M., Hanekamp, M., Cremers, A. H. M., & Neerincx, M. A. (2013). *User Requirement Analysis of Social Conventions Learning Applications for Non-Natives and Low-Literates*. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 8019 LNAI (Issue PART 1, pp. 354–363). Springer.