MALMÖ UNIVERSITY

Technology and Society

Exploring Emely: An exploratory case study on the usability and user experience of a conversational agent for L2 learning

Utforskning av Emely: En explorativ fallstudie om användbarhet och användarupplevelse av en konversationsagent för andraspråksinlärning

Julia Ahrling
Jonna Franzén

Degree: Bachelor
Main field: Computer and information science
Program: Information architect
Date of final seminar: 2023-06-02

Supervisor: Nancy Russo
Examiner: Sally Bagheri

Abstract

This study focuses on evaluating and enhancing the user experience of Emely, a conversational agent aimed at improving language skills for second language learners, particularly those who want to increase their chances of securing employment in Sweden. Usability testing was conducted in two test rounds, with the first round providing design implications for the user interface in the second round. However, assessing the effectiveness of the interface improvements was challenging due to low Swedish proficiency among the test groups consisting of potential users of Emely. Although the study did not result in design implications for the user interface, important findings highlight the need to adapt conversational agents, like Emely, for users with low literacy levels and illiteracy, emphasizing the importance of inclusive design for effective language learning support.

Keywords: conversational agents, L2 learning, usability, user experience

Sammanfattning

Denna studie fokuserar på att utvärdera och förbättra användarupplevelsen av Emely, en konversationsagent som syftar till att förbättra språkfärdigheterna hos andraspråksinlärare, särskilt de som vill öka sina chanser att få anställning i Sverige. Användbarhetstester genomfördes i två omgångar, där den första omgången gav designimplikationer för användargränssnittet i den andra omgången. Dock var det utmanande att bedöma effektiviteten av ändringarna i gränssnittet på grund av låg svenska språkkompetens bland testgrupperna. Även om studien inte resulterade i bekräftade designimplikationer för användargränssnittet, framhäver fynden i studien behovet av att anpassa konversationsagenter för andraspråksinlärning, såsom Emely, för användare som är analfabeter eller har en låg läskunnighet. Det betonar vikten av inkluderande design för hjälpmedel som Emely ska kunna vara ett effektivt stöd för språkinlärning hos sin målgrupp.

Nyckelord: konversationsagenter, andraspråksinlärning, användbarhet, användarupplevelse

Table of Contents

1. Introduction

Statistics Sweden (Swedish: Statistiska Centralbyrån) provides insights into the labor market and unemployment in Sweden. In a recent report, focusing on the disparities between native-born and foreign-born individuals, Statistics Sweden highlights the significant impact of factors such as education level and parental educational background on employment opportunities. Moreover, language skills emerge as a critical determinant of employment prospects, with proficiency in the Swedish language being a key factor [1]. There is a lot that comes into play when it comes to learning a second language such as vocabulary, sentence structure and how to stress words but things like culture and context also become important [2], [3].

One area that has seen a growing popularity in recent years is the use of conversational agents for second language (L2) learning [4]. These agents leverage advanced technologies, such as machine learning, deep learning, and natural language processing, to engage in natural conversations with learners and help them practice their L2 skills [2]. However, to ensure the success of any digital artifact, good user experience and usability are crucial factors [5]. Meaning that users need to have a positive experience while interacting with the conversational agent, finding it easy to use and effective in helping them improve their language skills.

In this context, Emely, a virtual second language assistant, has been developed to provide opportunities for L2 learning with a big focus on conversational practice [6]. In this thesis, we evaluate the overall user experience and usability of Emely, to identify areas of improvement for enhancing the overall usability and effectiveness of the virtual assistant. This evaluation will contribute to the development of more user-friendly and effective language learning tools, addressing the specific needs of the intended target group.

Based on our findings, we seek to provide design implications, including aspects such as navigation, interaction design, feedback mechanisms, and language prompts, if needed to ensure a seamless and engaging user experience. These design implications will help inform future iterations of Emely, as well as the design and development of similar tools, with the ultimate goal of creating more effective and user-friendly tools for supporting language learning among foreign-born individuals in Sweden and beyond.

1.1. Purpose and research questions

In this study, we focus on improving the perceived user experience and usability of the conversational agent Emely for the intended target group being foreign-borned individuals living in Sweden wanting to improve their conversational skills in Swedish with the main purpose of securing employment. As mentioned earlier, Emely is a conversational agent designed to facilitate second language acquisition, particularly in the conversational part of learning a second language [6] since it has a significant role in

searching and getting a job [1]. However, the user experience and usability of Emely from the user's perspective remains unknown. Although there is some research on the design of conversational agents for L2 learning when it comes to the system itself, there is not much research that focuses on the user experience and user interface of these.

To address this gap, we conducted a case study to identify the key challenges and limitations in the current user interface of Emely. The data was collected through observations in a usability test with users from the intended target group followed by semi structured interviews and a questionnaire. The goal was to identify design implications for improving the usability of Emely and inform the design of an improved version of the conversational agent.

1.1.1. Research questions

This study aims to contribute to the growing body of literature on the user experience with focus on usability of digital artifacts, particularly conversational agents. Our findings will provide insights into the design implications for improving the usability of conversational agents, especially for individuals wishing to improve their second language skills in order to better integrate in society. The research questions to be answered in this report in order to address the gap mentioned above are the following:

What are the design implications for improving the perceived user experience and usability of the conversational agent Emely?

- What are the key challenges and limitations in the current user interface of Emely as perceived by the intended target group?
- How can the user interface of Emely be improved to enhance the efficiency and satisfaction of its use for the intended target group?

2. Background

In this section, the background of the topic is presented, providing an overview of its significance and context within the field. The purpose of the research and the research questions are also presented. At last, a review of relevant literature, identifies research gaps, and outlines the objectives of the study is presented.

2.1. Importance of learning Swedish as a second language

Foreign-born individuals in Sweden often face language-related barriers when seeking suitable employment. A higher proportion of foreign-born individuals in Sweden, particularly foreign-born women, identify language proficiency as a significant obstacle compared to their counterparts in the EU [1]. This highlights the pressing need for foreign-born individuals to acquire proficiency in Swedish to overcome these challenges.

Learning Swedish is instrumental in addressing language-related barriers, facilitating effective communication in the workplace, and fostering a better understanding of the local culture and work environment. It is not only about language proficiency but also about promoting successful integration into the Swedish labor market.

Recognizing the significance of language acquisition, Sweden has implemented various initiatives and programs to support foreign-born individuals in learning Swedish. These include language courses, vocational training programs, and integration initiatives aimed at equipping individuals with the necessary language skills for active participation in the labor market [7].

2.2. Conversational agents

Conversational artificial intelligence (AI) and conversational agents are often used interchangeably to refer to computer systems that engage in natural language conversations with humans. Conversational AI is a broader term that encompasses the field of artificial intelligence technologies used to enable computers to understand and respond to human language in a conversational manner. Conversational agents, on the other hand, specifically refer to the applications or programs that use conversational AI technologies to engage in conversations with users [8], [9, Ch. 1], [10].

Chatbots, virtual assistants, and customer service agents are examples of conversational agents that utilize conversational AI to simulate human-like conversations and provide information, assistance, or other services. These systems utilize natural language processing (NLP) and machine learning (ML) to interpret and generate human language, making it possible for users to interact with technology through conversation for various tasks. Conversational AI and conversational agents aim to create a seamless and intuitive communication experience between humans and computers, enabling more convenient and interactive interactions with technology [9, Ch. 1], [10].

In recent years, conversational agents have gained popularity as a tool for various purposes in different industries, ranging from customer service and support to virtual assistants for productivity and scheduling, as well as language learning and tutoring. The versatility of conversational agents has led to their integration in diverse sectors such as healthcare, finance, e-commerce, entertainment, and also in the field of education [4], [11], [12].

2.2.1. CA's for second language learning

One particular area of education where conversational agents are increasingly being developed for and used is in second language (L2) learning [3], [4]. These agents use advanced technologies mentioned above such as machine learning, deep learning, and natural language processing to engage in natural conversations with learners to help them practice their L2 skills [13].

In order to learn a new language there are multiple factors that come into play such as having to learn a new vocabulary and how to build sentences, cultural context, pronunciations and actively practicing conversational skills [4]. It is also argued that internal factors such as motivation, study habits, attitude, intelligence, self esteem together with external factors such as environment and access to learning material, to mention a few, play a role when it comes to who succeeds in learning a new language [3].

When it comes to self esteem in second language learning, studies show that some L2 learners prefer to practice their conversational skills with a conversational agent instead of other people because the agents provide a flexible and accessible way for learners to practice their language skills anytime, anywhere without being judged [2]. Instead, they can be at home in their natural setting instead of being in a classroom environment which can cause anxiety for some individuals [14, pp. 162-163].

Even though conversational agents for L2 learning hold great promise for those seeking to improve their skills, there are also challenges. Some of these challenges are the current limitations in technology used to develop these tools, leading to users experiencing unnatural conversations characterized by robotic voices and encountering failures when providing incomplete sentences or receiving illogical answers [4].

2.3. User experience

User experience, UX, is a word that is frequently mentioned in many contexts. According to international standards on “Ergonomics of human-system interaction: Human-centered design for interactive systems”, ISO 9241-210:2019 [15], user experience is a “user’s perceptions and responses that result from the use and/or anticipated use of a system, product or service” and the users’ perception “include the users’ emotions, beliefs, preferences, perceptions, comfort, behaviours, and accomplishments that occur before, during and after use.” [15].

Experts within the field [16] defined three key characteristics of user experience that include user involvement, interaction with a product or system, and the interest in observing or measuring the user's experience. Without user behavior or potential behavior, it's simply a measure of attitudes or preferences. Any system or product, digital or physical, can be evaluated from a user experience perspective as long as there is some type of interface between the product or system and the user making the interaction possible [16, p. 4], [15].

2.3.1. Usability

According to international standards on “Ergonomics of human-system interaction: Usability”, ISO 9241-11:2018 [17], usability is the “extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use”. In the context of interaction between a user and a product effectiveness refers to the amount of accuracy a user completes a task with, efficiency refers to the amount of effort it takes for the user to

complete a task and satisfaction refers to how well the user experience during the tasks meet the user's needs and goals [17], [16].

Usability testing is a technique used in user experience (UX) design to evaluate how easy or difficult it is for users to interact with a product or service. It involves observing and collecting feedback from users as they try to perform tasks using a system or product. The purpose of usability testing is to identify areas where users may encounter difficulties, confusion or frustration when using a product. By observing user behavior and collecting feedback, it is possible to gain insights on how to improve a product's design and functionality, which leads to a better user experience [18].

2.3.2. UX metrics

To actually measure user experience in order to evaluate and improve it, there has to be a set of measurements for doing so, therefore a set of UX metrics have been developed [16, Ch. 1]. Metrics is a well known concept that exists everywhere in our daily life as we measure things like the volume and weight of ingredients for making our dinner, or the speed of the car we are driving to prevent us from going too fast. The important thing with metrics is that they stay the same, regardless of who measures or what is being measured, and that they are being collected in a consistent way. Examples of UX metrics are task success that can tell whether the user completed a given task or not, task time that can tell in what time range the user completed a given task or efficiency that can tell the amount of effort required by the user to complete a given task [16, Ch. 1].

2.4. User interface design

As mentioned above, user experience design is concerned with designing the overall experience that a user has while interacting with a digital product, including the user's emotions, attitudes, perceptions, and behaviors. While UX refers to understanding the user's needs and goals, and designing products in a way that makes them intuitive and easy to use in order to meet those needs and goals, user interface (UI) design, on the other hand, is concerned with the visual and interactive elements of the digital product, including its layout, typography, color scheme, and graphical elements. UI design is therefore about how to create an attractive, engaging, and consistent visual language that supports the user's interaction with the product. However, both are important for creating successful digital products that meet the needs and expectations of their users [19].

2.5. Emely

Mastering the Swedish language in both speech and writing is important for integration into Swedish society. Unfortunately, many foreign-born individuals or individuals with foreign-born parents have limited language skills, which limits their opportunities to obtain employment or access societal information. According to the Swedish Statistical Bureau's labor force surveys, the unemployment rate among foreign-born individuals was 18.4 percent in 2021, compared to 4.4 percent among Swedish-born individuals [1].

To assist individuals who need to improve their language skills in Swedish, with a particular focus on conversation and dialogue, the language assistant Emely has been developed by a company named NordAxon. Emely is a conversational agent that provides users with a realistic and meaningful learning experience by allowing them to practice their Swedish language and communication skills in various dialogue situations in a comfortable and interactive way. NordAxon also hopes that regular use of Emely can also help users improve their social skills and overcome anxiety and social phobia associated with limited communication skills [6].

Emely is a web-based conversational agent, which means that the user interacts with a graphical user interface (GUI) in order to perform tasks such as creating an account, log in, choosing the type of conversation to have. However, the conversation part of Emely is a conversational user interface where natural language processing (NLP) enables the user to interact with her using spoken or written language and then simulating a human-like conversation. NLP techniques allow Emely to interpret and understand the user input, generate appropriate responses and provide relevant answers based on the user input and context. Emely is not what you call an embodied conversational agent, meaning that the web application does not have an animated avatar with facial expressions and gestures as feedback. However, she is represented by an icon giving the users a visual representation of her.

The main feature of Emely is, as mentioned above, the ability to have a human-like conversation to practice the Swedish language, and she has three different modes in order to best meet the users' needs.

1. **Job interview** - makes it possible for the users to practice their swedish skills as well as interview skills in several different domains
2. **Taking a coffee** - making it possible for the users to practice small talk and more informal Swedish
3. **Easy swedish** - enables those with a lower language level to talk to Emely but with fewer words and not as long a conversation

2.5.1. User interface

Figure 1. Home page

Figure 2. Conversation view

2.6. Limitations

The limitations of the study are primarily due to the consultation with NordAxon, who want to focus on improving the user experience for the target group. This has led to narrow scope and concentrate solely on identifying the elements that need to be improved to enhance the overall user experience, usability and user interface. Due to time constraints, we cannot investigate all the factors that contribute to a better user experience such as language improvement and conversational flow.

In the context of the study, the main functionality refers to the key features and functions that are essential to its purpose and overall user experience of Emely's user interface. These features will be analyzed and evaluated to gain an understanding of how they can be improved to enhance the user experience for the target group.

On the other hand, "individual aspects" refer to specific, isolated components of Emely that may contribute to the overall user experience. Navigation on the page is an example of an individual aspect, which may affect the user experience, but is not essential to Emely's primary purpose. As a result, the study will not conduct an in-depth analysis of such components. It's worth noting that the decision not to analyze individual aspects in depth is not due to their lack of importance, but rather a result of

the study's limitations and focus on the primary objective of improving the user experience for the target group within the given time constraints.

Another thing worth mentioning is that the usability tests were carried out in the premises of the labor market administration in Helsingborg, because it is an environment where Emely is used. However, once the user has an account for using Emely they can use the conversational agent everywhere. This study has not tested Emely in other environments such as the users homes.

2.7. Previous research

For this study, researching the fields of conversational agent design when it comes to user experience, user interface and usability together with conversational agents in second language learning has been of high interest. To get an overview of the research within these fields, and potential parts missing, a literature review was conducted.

2.7.1. Conversational agent design

Several studies have emphasized the critical role of chatbot conversations in UX. A study found that the quality of chatbot conversations is essential for a good user experience [20]. According to the study made by Følstad et al. [21], incorrect response from the chatbot does not necessarily result in a negative experience, as long as the chatbot provides a simple route for further communication with human customer service agents.Users value conversational intelligence and the chatbot's ability to retain context. Additionally, message interactivity, conversational flow, and repair are crucial for a positive experience. Adapting the conversation to the user type is also essential [22, 23].

Other studies have addressed the relationship between the appearance and personality of conversational user interfaces and UX. Murgia et al. [24] found that users' perceptions of the chatbot depend on its self-presentation as human or machine. Thies et al. [25] and Smestad and Volden [26] found that users prefer socially oriented or engaging chatbot personalities over those that are more neutral or information-oriented. Finally, de Visser et al. [27] showed that conversational agents that are more human-like may have higher trust resilience than less human-like agents, meaning that they may be more trusted by users in situations characterized by deteriorating reliability in the information provided by the agent.

Several studies [28], [29], [30] also talk about the system architectural design of conversational agents and conversational user interfaces but none of them mention any implications on how the graphical user interface of web based CA’s can be designed to support the architectural design, which suggests that this could be an area to do more research on since these agents can not be used by the users without a user interface.

2.7.2. Second language learning

The willingness to communicate (WTC) is an important factor that can influence the success of second language learning. Previous studies have shown that learners who have a high level of WTC tend to be more successful in learning a new language, as they are more likely to seek out opportunities to use the language and receive feedback from others. In contrast, learners who have a low level of WTC may be less likely to engage in communication, which may hinder their language development [31].

An empirical study by Huang et al. [4] the use of chatbots in language learning within higher education was examined. Researchers found that chatbots were predominantly used in higher education due to the growth of online learning. The study highlighted the technological affordances of chatbots, including their ability to provide timely and personalized responses, which enhanced students' communication in target languages.

The study [4] identified three challenges with chatbots in language learning: technological limitations, novelty effects, and cognitive load. To address these challenges, researchers suggested that teachers play a leadership role in utilizing chatbots effectively, workshops to prepare students prior to the first use of the chatbot, and integrating certain principles to present the texts in the chatbots in the form of a conversation. Future research was recommended to explore the benefits of chatbots across different education levels, conduct longitudinal studies, employ objective measurements, compare chatbots with other tools, and investigate teachers' perceptions.

3. Method

This section describes the research method used in this study. There is also a description of the choice of data collection methods, as well as the metrics and tools used to collect data. The section also includes a discussion of alternative methods that was considered but rejected.

3.1. Case study

The chosen research approach for this study is a case study, which involves conducting a detailed investigation of Emely as a virtual assistant for second language learning, specifically focusing on the perceived level of usability from users' perspectives. Case studies are a widely used research approach that involves conducting a detailed investigation of a particular case or situation [32, Ch. 10].

While case studies mostly use quantitative data collection methods, this study uses a mixed method approach to gather both qualitative and quantitative data to gain a comprehensive understanding of the user experience with Emely. This approach allows for generating rich, qualitative insights from observations during the usability test sessions followed by in-depth interviews. Through these qualitative methods, users' attitudes, perceptions, and behaviors in relation to Emely's user interface design and

their experiences with the virtual assistant in the context of second language learning can be explored..

Furthermore, incorporating quantitative measures such as questionnaires enables the collection of quantitative data, which provides objective measures of usability. By utilizing scales like the System Usability Scale (SUS) and Net Promoter Score (NPS), quantifiable data on users' satisfaction and likelihood of recommending Emely can be obtained.

By combining qualitative and quantitative data collection techniques, a more comprehensive understanding of Emely's usability can be achieved, addressing the research questions and providing a detailed and nuanced analysis of the user experience.

3.2. Nordaxon

In the beginning of this study the researchers had several meetings with employees at NordAxon who have been, and are currently, developing Emely to get more details about the background and purpose of the development as well as the technical part of the product meaning which techniques have been used in the development.

Discussions ranging from what the users feel about Emely’s avatar/icon to what the users would feel about gamification features were held. Once the literature review was completed and the research focus was established in collaboration with NordAxon, the decision was made to evaluate the user experience and usability of Emely. During the development of Emely, test sessions were conducted and documented by NordAxon, which provided a basis for documentation analysis. However, the test sessions were unstructured and all test participants were in the same room using computers, resulting in limited notes and observations.

The evaluation sessions revealed the importance of optimizing the registration process. Many users encountered difficulties during this stage, indicating a need for simplification and streamlining to improve the overall user experience. Furthermore, the evaluation sessions revealed that it is important for users to be able to vary the speed of Emely's speech. Allowing users to adjust the speed of Emely's responses will cater to their individual preferences and comprehension abilities, ensuring optimal understanding and engagement.

3.3. Participant selection

The selection of participants for this study was partly based on the fact that the development of Emely has not yet been officially released. AMF Helsingborg provided a test group consisting of individuals who are intended users of Emely, some of whom have already had prior experience with Emely. By utilizing this test group, it aimed to gather insights and feedback from individuals who are representative of the actual users.

3.4. Usability test

Emely has been developed by NordAxon in cooperation with The City of Helsingborg and The Labor Market Administration there, whose goal is that the residents of Helsingborg should be able to find employment and support for themselves. As explained earlier, Emely is able to support people with Swedish as a second language in interviewing for jobs, which suits the goals of The Labor Market Administration. In order to carry out the usability test, a test group has been used in two separate test sessions.

The design of the test was based on the primary goal of the study, to answer the research questions. In order to find out the perceived user experience and usability of Emely, seven tasks were created that represent the main functionality of the system. The seven tasks included creating an account, engaging in a conversation with Emely, finding the help and suggestion buttons, locating statistics on progress, adjusting settings for voice speed, sound quality, and native language, and logging out. Each task is assessed with metrics such as either success or fail, as well as how long it took the participant to complete the task.

Usually a traditional moderated usability test will take place in a lab setting [16, p. 53], however, the tests were carried out in the premises of the labor market administration in Helsingborg, because it is an environment where Emely is used, which the researchers considered to be an advantage in order to get a better overall picture of what conditions exist for the use of Emely. The tests were conducted in a room with one participant at a time, where each participant completed the series of seven tasks on either a mobile phone, computer or tablet, depending on the participant's preference. The user testing sessions were led by one researcher who assumed the role of the moderator, taking charge of providing participants with tasks and offering assistance if necessary. In contrast, the other researcher took the role of an observer, adopting a passive stance primarily focused on observing and documenting the participants' interactions without interfering or providing guidance.

Before the tests were carried out, careful planning was done to ensure that the results would be as good and objective as possible. In order to ensure this a test script was developed containing the necessary information that had to be given to the test group before starting the tests (see Appendix A). The test group was informed that the primary objective was to evaluate the usability and design of the product, rather than assessing their individual performance. They were also informed that the results would be anonymous and each participant was asked for consent when it came to screen and audio recording. The script also contained a part that was only for the moderator of the test including the tasks and the definitions of success and failure for each of the tasks [16]. It should be noted that the moderator and the observer were aware of the participants' limited Swedish language proficiency prior to the test, and this was taken into account during the planning phase. Instead of only having only Success or Fail, there is also a category of Success with Guidance [16, pp.71-72]. This category indicates that the participant was able to complete the task with the assistance of the moderator.

After the general information about the test was given to the participants they were given test instructions. Each task was written down in Swedish on a separate note and given one at a time to the participants and all the tasks were done in the exact same order in all tests. The participants were informed that when they considered themselves to have succeeded with a task they would say something like “done”, this to avoid the moderator mistakenly assuming that the participant completed the task, when in fact the participant may have completed the task by mistake. The participants were informed that they should try to succeed with the task by themselves but in case they considered it impossible, they could ask the moderator for guidance.

3.4.1. Metrics

Time on task is used in usability testing to measure the efficiency of a product or system, in other words the amount of effort it takes for a user to succeed with a task. Time on task is also important to measure for products intended to be used on a regular basis and where the tasks are being repeated by the user which is the case with Emely [16, p. 75], [17]. Time on task can be hard to measure consistently if no definition of when the task starts and ends has been stated. To avoid this, the researchers decided that the start will be as soon as the participant tells the moderator that they understood the task and feels ready to begin. The end of each task was when the participant said that they were done.

Task success is said to be the most commonly used usability metric and is used for measuring how well users succeed to finish tasks, usually representing key pieces of information or functionality in a product or system. This data can give valuable information on how effective a digital artifact is to use and if there is a big number of users who fail the tasks, it’s an indication that these things need to be taken a closer look at [16, p. 65]. In the case of this study, task success was chosen in order to see if the intended users were able to succeed with the tasks representing the key functionality of Emely. If not, that could be an indication that the system is not as effective as it should be to meet the users goals and needs.

3.4.2. Observations

Observations are commonly used in case studies, as well as in usability tests, as a qualitative data collection method to gain understanding about what and why users think the way they do about, in this case, a digital artifact. According to Oates [32, p. 202] “Observing is something most of us do a lot of the time: seeing, hearing, noting, analysing, forming theories, making inferences, imposing meaning.” which is what this study aims to do when it comes to the user experience and usability of Emely.

During the usability test described above, one of the researchers observed the participants’, while the moderator noted the verbal and nonverbal expressions, feelings, and behavior during each task. To avoid missing any relevant information, the researchers used screen recording with audio during the first test session. This made it

possible to review and analyze how the user navigates and interacts with the product at a later time.

3.5. Interviews

Interviews are also commonly used in both case studies [32, p.187] as well as in usability studies, especially summative ones on a completed product like in the case of this study 33, p.197], to provide the developers with deeper insights that can be used in future iterations of the development. It is a suitable data collection method when the researcher aims to obtain detailed information, ask open ended questions and explore emotions, experiences or feelings that cannot be observed, or described as easily in a questionnaire with predefined responses.

There are several different types of interviews, but the best for this particular study were considered by the researchers to be semi-structured interviews. The researchers wanted a predefined set of questions to make sure that a certain consistency through the interviews were held but also in order to make sure that answers were collected on the data considered most important to collect to answer the research questions. However, the researchers wanted the opportunity to ask other questions based on the participants' answers and be able to ask the questions in the order that felt most natural, as well as have a more relaxed setting [32, p. 188].The questions that the researchers chose to have predefined were the following:

Q1: Have you used Emely before, or is it the first time?
- If yes, do you use Emely in your spare time/at home?

Q2: Do you prefer talking or writing with Emely?
- Why do you prefer talking/writing?

Q3: What were your feelings about using Emely? Was it fun, boring?
Q4: Was there something particularly easy/hard while using Emely?
- If something was hard, what could be improved to prevent this?

Q5: What was the best thing about using Emely?
Q6: What do you think about Emelys avatar/icon, how she looks?

3.6. Questionnaires

Post session ratings are used in usability testing after the interaction with the product or system being evaluated is done in order to collect data on the perceived usability from the participants perspective [16, p. 137]. Using the metrics below can help identify specific areas where a product may need improvement and track progress over time to ensure the user experience is constantly improving using them as benchmarks. The SUS and NPS questionnaire can be found in Appendix B.

3.6.1. SUS

The System Usability Scale (SUS) is a commonly used tool that measures the perceived usability of a product. It's a standardized questionnaire that consists of a 1-5 scale that

assesses users' perceptions of the product's ease of use, learnability, efficiency, satisfaction, and overall usability where half of the statements are worded positively and half negatively [16, p.137]. There are 10 questions in total.

SUS is considered a quick and reliable tool for gathering feedback from users, and it can be administered at different stages of the design process to identify potential areas for improvement. Its popularity in usability testing is because of its simplicity, ease of administration, and ability to provide a standardized measure of user satisfaction and usability across different systems and contexts [34]. Even though the questions were taken from a standard SUS questionnaire, it should be noted that slight moderations in the language were made to make it easier for the participants to understand.

3.6.2. NPS

Net Promoter Score (NPS) is a customer loyalty metric that measures the likelihood of customers recommending a company, product, or service to others [35]. The NPS score can help organizations understand how well their products or services are meeting customer needs and expectations, in other words by asking customers how likely they are to recommend a product or service, organizations can gain valuable insights into the usability and overall user experience.

Customers who are highly satisfied with the usability of a product or service are more likely to become promoters and recommend it to others, while those who are dissatisfied are more likely to become detractors and discourage others from using it. NPS provides a simple, standardized way to measure customer satisfaction and loyalty, and can be used to track changes in customer satisfaction over time and identify areas for improvement.

NPS is calculated based on a single survey question that asks customers how likely they are to recommend a product or service to others on a scale of 0 to 10. Respondents are then categorized into three groups: detractors (0-6), passives (7-8), and promoters (9-10) [35].

3.7. Methodology discussion

This section will discuss the chosen research method and data collection method, as well as provide an overview of alternative methods and data collection approaches that were considered but ultimately not used in the study.

3.7.1. Research method

The researchers discussed field study as an alternative method. While both case studies and field studies are valuable research methods, each has its own strengths and limitations depending on the research question and context.

A case study allows for in-depth exploration of a specific example or instance of a phenomenon, such as Emely, which can provide more detailed insights than a broader field study. With a case study, it can focus on a specific aspect of Emely's user interface design and explore its impact on the user experience in depth. In contrast, a field study would require collecting data on a larger group of L2 conversational agents, which may make it more difficult to delve into the specific nuances of Emely's design. It also allows control of certain variables and conditions, which can be difficult to do in a field study. For example, with select participants who have similar levels of language proficiency, which can help to isolate the effects of Emely's user interface design on the user experience. In a field study, it may be more challenging to control for these variables, which can make it harder to draw definitive conclusions about the impact of user interface design.

Finally, a case study allows for more detailed data collection methods, such as in-depth interviews and observations, which can provide a more complete picture of the user experience. By focusing on Emely and using qualitative research methods, it can generate detailed insights into the impact of user interface design on the user experience that can inform the design of future conversational agents for language learning.

3.7.2. Data collection methods

The use of the Think Aloud method is considered as an effective way to gather data on users' thoughts about a design [16, Ch. 3]. Nielsen quotes "Thinking aloud may be the single most valuable usability engineering method." [36, p. 195]. It has many advantages, such as being cheap, robust, flexible, convincing, and easy to learn. However, there are also potential risks, such as an unnatural situation, filtered statements, and biased user behavior. Most people want to appear intelligent and will not speak until they have carefully thought through the situation [37].

Therefore, the think-aloud method was not chosen in this study because of the target group. Given that they already have limited knowledge of the Swedish language, requiring them to express their feelings and thoughts could adversely affect the test, which would, in turn, impact both validity and reliability. It would also have been difficult to use think aloud, especially in the task where they had to have a conversation with Emely, because then they are already talking aloud to her. It would have been nearly impossible considering the risk that the participants' thoughts and feelings about the experience would have entered the dialogue by mistake, which would have negatively affected the results.

3.8. Reliability and validity

In the context of usability and user experience evaluation, reliability ensures that the measures used produce consistent results over time and across different evaluators. If the measures used are not reliable, the resulting data may not be consistent and may vary depending on the evaluator, making it difficult to draw reliable conclusions about the product's usability and user experience [38]. To enhance the reliability the chosen

research method and data collection methods are described and motivated in accordance with the study and its goals making it easy for readers or other researchers to gather their own opinions about them. The study also provides the necessary information and documents for making an exact replica of the study. Moreover, the participants of the study are representatives of the intended target group which enhances the reliability further.

Validity, on the other hand, refers to correctly measuring what it is intended to measure, and ensures that the metrics used accurately reflect the user's experience with the product, in this case Emely. If the measures used are not valid, the resulting data will not accurately reflect the user's actual experience, which can lead to incorrect conclusions about the product's usability and user experience [38].

4. First test round

In this section the results and analysis from the data collection from the first test round will be presented. The metrics from the user testing will be presented, as well as the data from the observations, interviews and the self reported metrics to get a SUS score and an NPS score. Furthermore, analysis of the results and proposed changes in the user interface of Emely based on the findings will be presented.

The test group for the first test round consisted of five people who had Swedish as their second language, all in some way participating in The Labor Market Administrations activities and they were all real potential users of Emely. The participants had diverse backgrounds and language abilities. Participant 1 (P1) and participant 5 (P5) had a higher level of proficiency in Swedish. However, P5 was unable to read or write in the native language. Participant 2 (P2) had an intermediate level of Swedish, and participant 3 (P3) and participant 4 (P4) had a low level of proficiency in Swedish.

4.1. Interface

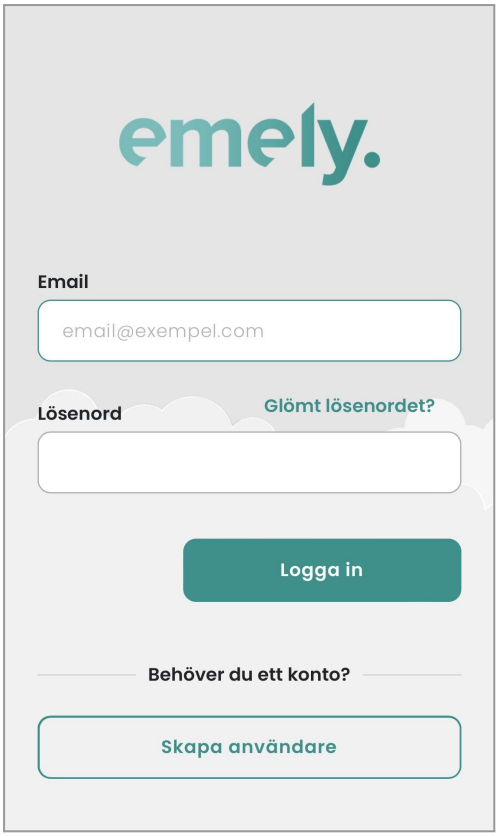


Figure 3. Log in / create account

Figure 4. Home page

Figure 5. Conversation flow

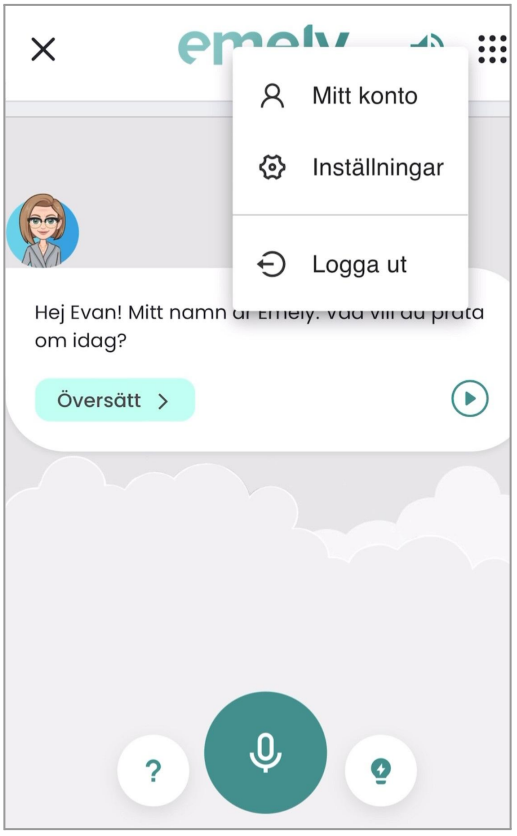


Figure 6. Menu options

4.2. Results

The tasks mentioned in this section, and described and motivated in section 2.2.2, refer to the following: 1. Create Account, 2. Have a conversation with Emely, 3. Find the help button, 4. Find the suggestion button, 5. Find statistics of progress, 6. Find settings for voice speed, sound quality and native language and 7. Log out.

4.2.1. Metrics

Time on task

Due to the fact that it was the first time the test was conducted and there was no clear criteria for what constituted a success or failure when it came to time on task and the tasks also vary slightly in difficulty, therefore the observer made the decision to not set a time limit for completing each task. The results of the tasks performed by the participants are presented in a table, where "P" represents the participant number and "Task 1," "Task 2," etc. refer to the different tasks that were assigned.

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
P1	5.58 min	19.53 min	1.12 min	8 sec	2.38 min	8 sec	8 sec
P2	2.07 min	13.58 min	3 sec	7 sec	34 sec	43 sec	5 sec
P3	4.21 min	16.20 min	37 sec	23 sec	3.05 min	11 sec	51 sec
P4	5.11 min	13.51 min	-	17 sec	-	-	-
P5	6.33 min	10.45 min	2 sec	53 sec	1 min	53 sec	3 sec

Figure 7. Time spent on each task.

P4 encountered difficulties in completing several tasks. Specifically, P4 failed to accomplish tasks 3, 5, 6, and 7. This is due to the fact that even though P4 informed the moderator that the tasks were completed, P4 did not carry out the intended actions as required. More details are explained in section 3.4.1.

Task success

The results are presented in three different categories: **Green Success**, which represents whether the participant was able to complete the task successfully on their own, **Yellow Success**, which represent that the participant was able to complete the task with the help of the moderator, and **Red Fail**, which indicates that the participant was unable to complete the task.

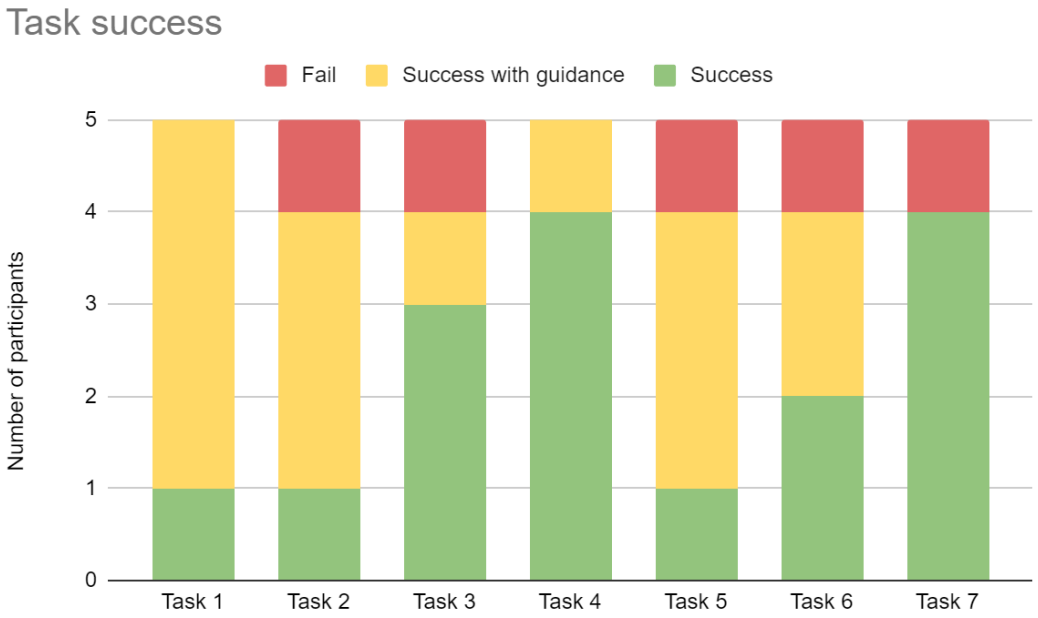


Figure 8. Task success among the participants of the first test round

4.2.2. Observations

Participant 1

During Task 1 (creating an account), P1 initially struggled to understand the instructions and only entered part of their email address as their username when trying to create an account. P1 also talked aloud throughout the task. During Task 2 (having a conversation with Emely), P1 was very happy and engaged in the conversation, asked Emely questions and expressed a desire to learn more from her. P1 encountered some technical difficulties but remained positive. During Task 3 (finding the help button), P1 had difficulty locating it and needed assistance. During Task 4 (finding the suggestion button), P1 found it quickly. During Task 5 (finding progress statistics), P1 was initially uncertain but persisted with help and encouragement. Task 6 (finding voice settings) and Task 7 (logging out) were both completed easily and quickly.

Participant 2

During the tasks, P2 was generally proactive and had a higher level of Swedish proficiency. P2 was also friendly and willing to help, offering to teach Emely Arabic. P2 was able to quickly complete most tasks, but struggled with finding the statistics on progress. Overall, P2 completed the tasks in a reasonable amount of time and without any major issues.

Participant 3

Participant 3 encountered difficulty in completing certain tasks such as creating an account, and required assistance in locating buttons and navigating the interface. P3's struggle with the language barrier was evident, necessitating aid in understanding

English instructions and translating them to Arabic. In the conversation task, P3 showed a lower level of language proficiency, necessitating assistance with each answer. The participant frequently forgot to record the messages and required clarification from the observer to proceed with the conversation. Furthermore, P3 displayed code-switching behavior, alternating between Swedish and English. The process of locating progress statistics was time-consuming, prompting P3 to ask for assistance. However, P3 was able to promptly locate settings for voice speed, sound quality, and mother tongue.

Participant 4

Participant 4 was generally positive and seemed to like Emely during task 1, but didn't understand why P4 had to fill in the password again. In task 2, P4 struggled with understanding and became frustrated. P4 focused heavily on what was written and deleted text when it wasn't expected. P4 had a low level of Swedish and was easily distracted, pausing the task to talk about other things. For task 3 and task 5, P4 didn't seem to understand what P4 was supposed to do. In task 4 and task 6, P4 completed the tasks but was unsure if P4 succeeded and needed help. P4 could not complete the log out task.

Participant 5

Participant 5 appeared calm and focused during Task 1 (creating an account), but needed guidance on how to create a password and where to input their email address. P5 missed selecting their native language and agreeing to the terms and conditions at first. During Task 2 (having a conversation with Emely), P5 had some difficulty starting the conversation and required assistance in using Emely's features. P5 struggled with Emely's repetitive responses and asked if P5 could use Emely at home. P5 quickly found the help and suggestion buttons during Tasks 3 and 4, but took some time to locate the progress statistics and settings for voice speed, audio quality, and native language. P5 logged out of the system immediately without any problems.

4.2.3. Interviews

Question 1

The first question of the interview was if the participants have used Emely before. The result shows that P2 and P3 have never used her before, P1, P4, and P5 answered that they have used her before, of which P1 has used her only one time before, while P4, and P5 have used her several times at home.

Question 2

According to the responses of the participants to the question of why they prefer to talk or write with Emely, it appears that P1 finds it easier to talk, but also notes that if someone is not proficient in Swedish, writing may be easier. P2 prefers talking with Emely because it is faster, while P3 prefers talking in general. P4 answered that they are comfortable with both talking and writing, but that Emely has difficulty

understanding them when they speak. Finally, P5 prefers talking over writing as they find it easier.

Question 3

Regarding the participants' opinions about using Emely, there were mixed responses. P1 finds some of Emely's responses boring and suggests that they should be changed, but also acknowledges that some of the responses are funny. P2 expresses that using Emely is fun because she is always happy. P3 finds it enjoyable to talk about food with Emely. P4 only responds with "very good" and provides no additional feedback. Finally, P5 thinks using Emely is fun and also mentions that she is easy to use.

Question 4

When asked about what they found easy or difficult when using Emely, the responses were varied. P1 suggested that it would be easier if Emely could provide more information and engage in conversation with the user. P2 found it difficult to have unprepared conversations, while P3 was indecisive and did not provide specific examples of what they found easy or difficult. P4 did not find using Emely difficult and enjoyed practicing their Swedish speaking skills. Finally, P5 did not specifically answer the question and instead spoke about other things, potentially due to language barriers hindering their understanding of the question.

Question 5

When asked about what they thought was the best thing about Emely, the participants had different opinions. P1 mentioned that they appreciated the ability to get information and learn, as well as feeling safe talking to Emely who does not judge or laugh. They also described Emely as kind. P2 highlighted the opportunity to practice speaking with Emely, without feeling stressed or pressured. P3 also enjoyed talking to Emely as the best aspect of the experience. P4 valued the ability to speak with Emely, though they preferred talking to a human. Finally, P5 appreciated how Emely could speak a lot and respond quickly, making it easier to talk to her. They also noted that some people may not want to talk, making Emely a good option for language practice.

4.2.4. Questionnaires

SUS

After the participants completed the usability test, they were asked to answer a SUS questionnaire. The result shows an average score of 64. The questionnaire, a total of 10 questions, and separate answers to each question can be found in Appendix B.

NPS

After the usability test, the participants were asked if they would recommend Emily to someone else, such as family or friends. The scale ranges from 1-10, with 1 being “Not likely at all” and 10 being “Very likely”.

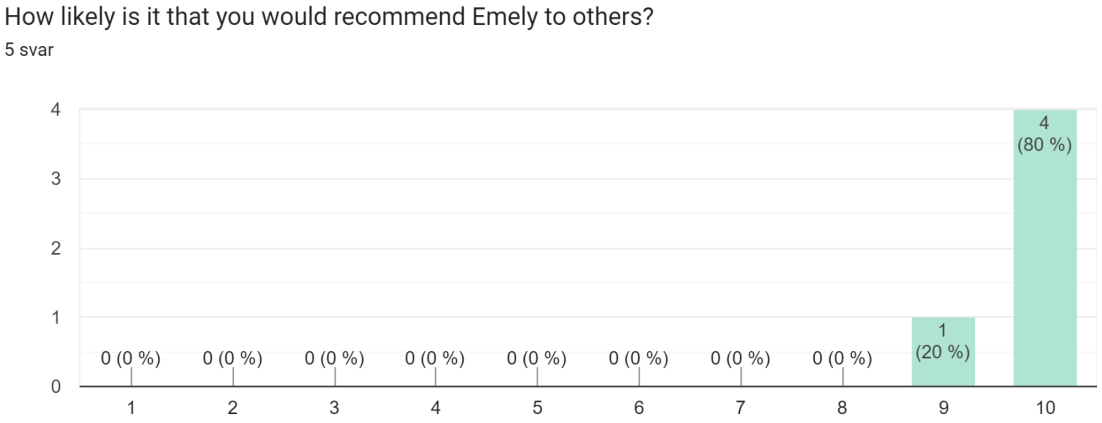


Figure 9. Results from NPS questionnaire (translated to English).

Figure 9 shows that out of the 5 participants, 4 responded 10 meaning that they were very likely to recommend Emely to others. 1 answered 9 meaning that they are likely to recommend Emely to others.

4.3. Analysis

Only one participant was able to complete the task of creating an account without any assistance, and this was the participant with the highest level of Swedish proficiency in the test group. Out of the five participants, four encountered difficulties while creating an account, with one participant entering only part of their email address as their username. The researchers noticed during the observations that when the participants were asked to create an account they immediately started to fill in the login form, which is on top of the landing page when not logged in, not noticing the text “Do you need an account?” followed by the button “Create user” (see Figure 3). This led to several errors trying to complete the task indicating that the target group could benefit from improvements in these steps.

The researchers also found that the majority, 4 out of 5, of the participants needed assistance in how to record and send messages to Emely when having a conversation with her. When entering a conversation Emely immediately starts to talk and 4 out of 5 started talking back to her without noticing her message in written text on the screen. However, once the moderator of the test explained how to record and send only 1 out of 5 continued experiencing problems with this which indicates that once the user figures out how to record and send a message it becomes less of a problem.

During the conversation with Emely, 2 out of 5 participants were able to engage in the conversation without guidance from the moderator meaning that they understood how to record their answers and send them to Emely. 3 of the 5 participants needed help understanding how to record and send the message in order to have the conversation. Out of these three, one required assistance for each separate interaction with Emely

during the conversation. Another one out of these three focused heavily on what was written when recording and receiving messages, in case the recording was wrong the person would delete the text and then try to record again or change the text by typing instead. This suggests that it needs to be more clear that the users need to press the microphone icon to start and stop the recording, and then send the message to Emely.

In Task 3, which involved finding the help button, 2 out of 5 participants were able to locate the button without any issues, while 2 participants required assistance, and 1 participant did not seem to understand the task. However, all participants were able to locate the suggestion button (Task 4) without any issues. During Task 5, which involved finding the progress statistics, 3 out of 5 participants experienced difficulty and required assistance, while 1 participant did not understand the task and found it challenging, and 1 participant eventually found the statistics after some time. In Task 6, which involved finding the voice settings, 2 out of 5 participants were able to locate the settings without any issues, while 2 participants required some time to find the settings. One participant completed the task but was unsure if they succeeded and required assistance. These findings all indicate that the target group could benefit from a different kind of navigation making it easier to find and access the functionality of Emely.

Finally, during Task 7, which involved logging out, 4 out of 5 participants were able to complete the task without any problems. The one participant who did not complete the task simply did not want to log out, the participant wanted to remain logged in and continue to use Emely at home. This all shows that the participants encountered challenges with various tasks and those with a lower level of Swedish proficiency seemed to struggle more. This could indicate that language barriers may impact the usability of conversational agents, particularly in the context of account creation and navigation, and that the graphical user interface (GUI) of Emely needs improvement to make it more user-friendly and intuitive.

Based on the previous research and the results of the interviews, we can analyze and interpret the participants' responses in terms of the critical role of chatbot conversations in UX. Studies cited [22], [23] in the previous research suggest that the quality of chatbot conversations is essential for a good user experience, and users value conversational intelligence and the chatbot's ability to retain context. In the interviews, the participants generally found Emely to be a helpful conversational agent for language learning, regardless of their preferred mode of communication. This suggests that Emely's conversational intelligence and ability to retain context may have contributed to the positive user experience reported by the participants.

The previous research also highlights the importance of adapting the conversation to the user type [22], [23]. In the interviews, the participants had different preferences for talking or writing with Emely, indicating the importance of providing users with multiple communication options to accommodate their individual preferences and needs.

4.4. Changes in the UI

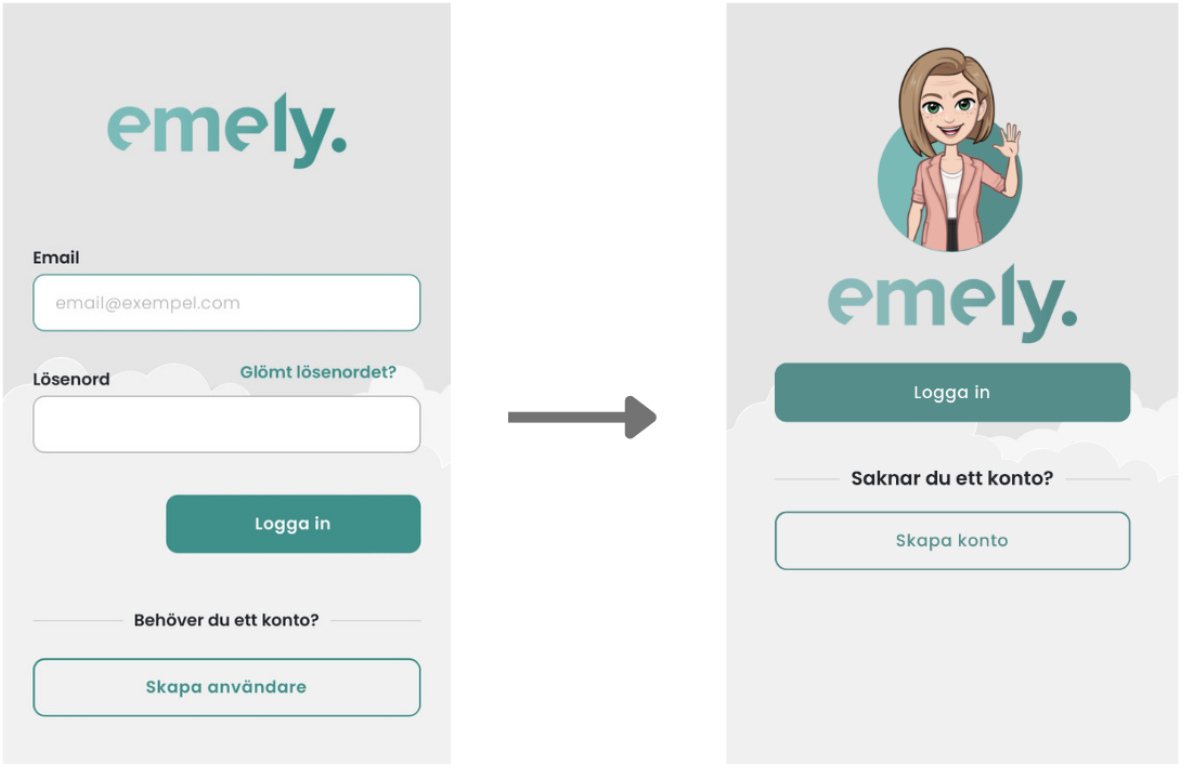


Figure 10. Changes in login/registration flow

As previously mentioned, 4 out of 5 participants had difficulties creating an account. The most common error the participants made in this step was that they immediately tried to fill in their email address in the form seen in the picture to the left in Figure 3. It is difficult to draw any conclusions and generalize as the test group only consisted of five people, but since problems with logging in also occurred during the testing NordAxon did during the development of Emely, it may be an indication that more users will have problems with the same thing in the future.

The researchers therefore took their cue from Nielsen's heuristics, which are a kind of rule of thumb for interaction design, and are widely used in the design and evaluation of digital products and services. When changes to the interface were to be made and the researchers were able to determine that one of his heuristics, error prevention, could be applied to the interface [39]. By removing the login form and instead only having a login button that leads to the form, the risk of the user making this particular error at this step is removed.

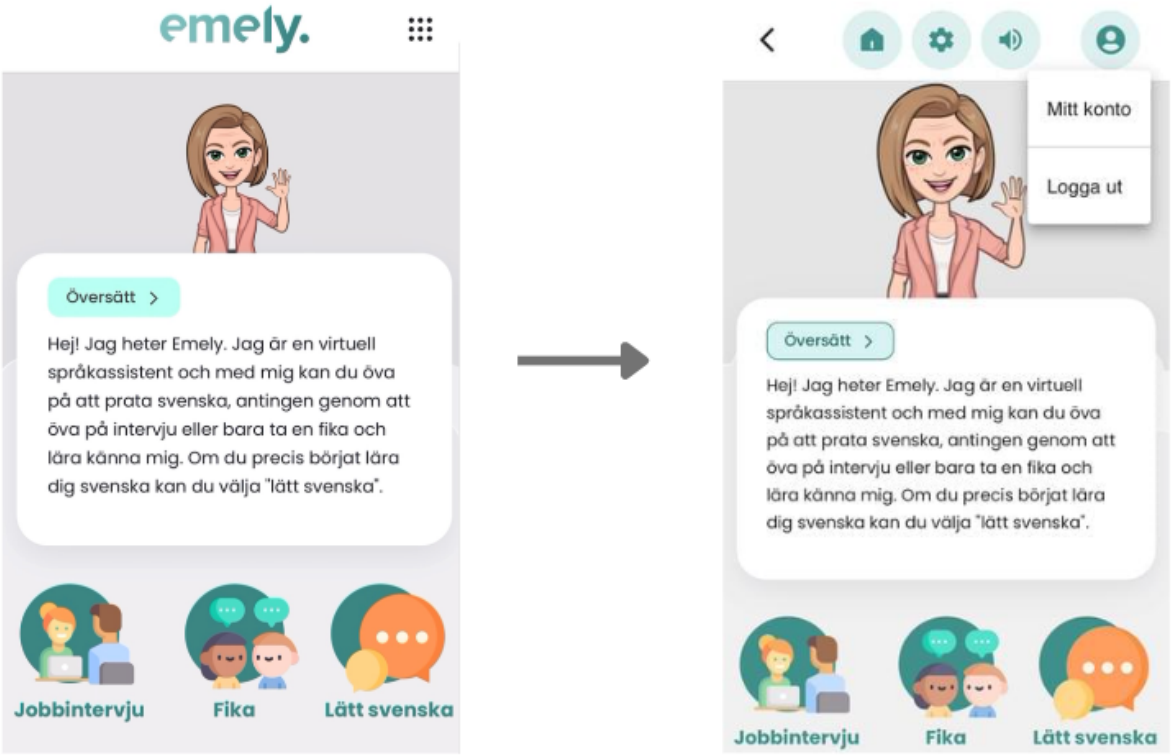


Figure 11. Changes in navigation/menu

The participants also struggled to find the basic functionalities hidden inside the navigation, the icon with the nine dots at the top right of the left image above. Therefore the choice to bring out the functionality from the menu and have them represented by icons instead where made.

5. Second test round

In this section the results from the data collection from the second test round will be presented. The metrics from the user testing will be presented, as well as the data from the observations, interviews and the self reported metrics to get a SUS score and an NPS score. Test session two aimed to replicate the setup of test session one, involving a test group of 5 individuals to evaluate Emely (regarding the same tasks, survey and interview) with the updated interface. However, test session two did not proceed as intended.

Similar to test session one, test session two took place at AMF Helsingborg, and the selection of participants remained unchanged, targeting individuals who were intended users of Emely. The main distinction in test session two was that the test group exhibited a significantly lower proficiency in the Swedish language. Participant 6 (P6) and participant 7 (P7) had an intermediate level of Swedish proficiency. Participant 8 (P8), on the other hand, had a low level of Swedish proficiency. Notably, both P7 and P8 mentioned their inability to read in their native language.

Despite the initial plan to include five participants in the test, the moderators decided to terminate the test session with only three participants, deeming the results unreliable. It is evident that the participants exhibited a limited understanding of the assigned tasks, as evidenced by their subsequent completion of questionnaires and interviews, which revealed a lack of understanding regarding the content of their responses.

5.1. Interface

In the version of Emely used in the second round of testing the suggested changes based on the results from the first round of testing were implemented. Pictures of this interface can be found in section 4.4.

5.2. Results

The tasks mentioned in this section, and described and motivated in section 2.2.2, refer to the following: 1. Create Account, 2. Have a conversation with Emely, 3. Find the help button, 4. Find the suggestion button, 5. Find statistics of progress, 6. Find settings for voice speed, sound quality and native language and 7. Log out.

5.2.1. Metrics

Time on task

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7
P6	4.43 min	15.08 min	17 sec	1.15 min	1.22 min	1.4 min	2.15 min
P7	3.55 min	13.19 min	1.46 min	57 sec	32 sec	38 sec	11 sec
P8	5.55 min	19.36 min	-	-	-	-	-

Figure 12. Time spent on each task

Figure 12 shows the time spent on each task in the second test round. Worth mentioning is that P8 started all tasks but gave up in trying to succeed Task 3-Task 7 almost immediately after being given the tasks, therefore there is no end time.

Task success

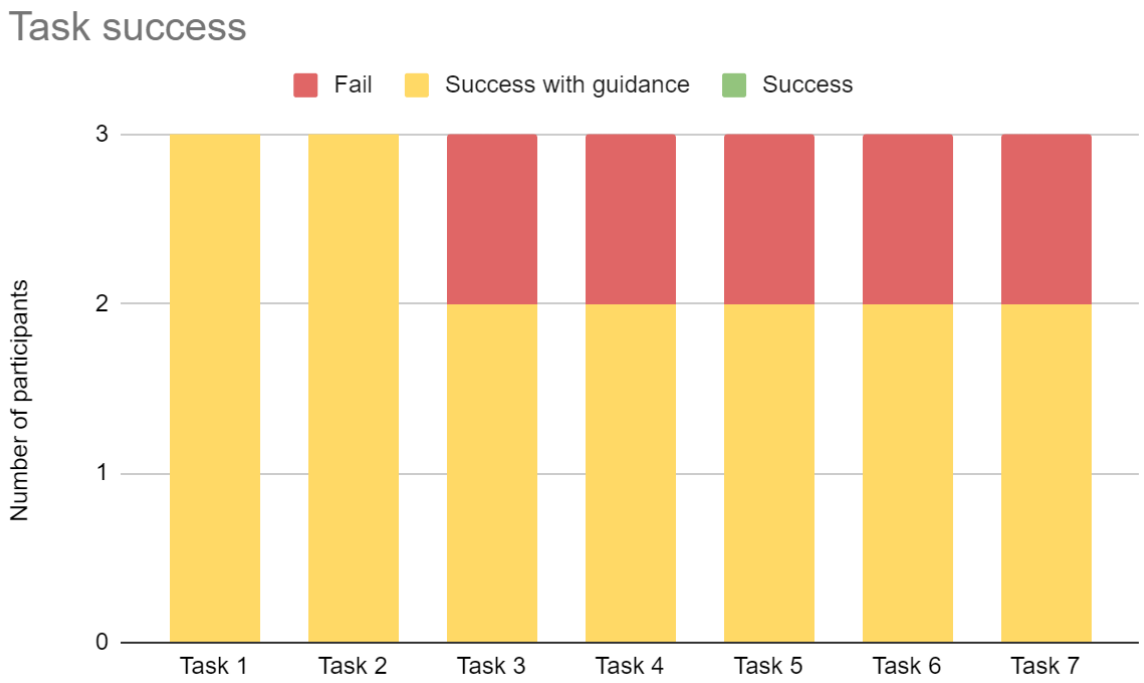


Figure 13. Task success among the participants of the first test round

Figure 13 shows that none of the three participants managed to perform any of the tasks without guidance. Worth mentioning is that P8 started all tasks but gave up in trying to succeed Task 3-Task 7, therefore those results count as failures.

5.2.2. Observations

Participant 6

During the account creation, P6 initially struggled to find the right options but eventually managed to fill in the required information and showed some confusion about creating a password.

During the conversation with Emely, P6 had some difficulty understanding instructions and needed assistance to send messages. P6 seemed relaxed and read responses more than listening. P6 occasionally translated responses into their native language. P6 gained confidence as the conversation progressed and stopped translating as frequently. P6 spoke loudly and expressed interest in dogs, while the user mentioned not wanting one. Emely misunderstood "no problem" as "ginger" due to the accent, leading to a humorous exchange.

In terms of finding buttons and settings, P6 had mixed results. P6 quickly found the help button, but needed more guidance to locate the suggestion button. P6 struggled to find progress statistics and eventually gave up, finding it too difficult. When it came to adjusting voice speed, sound quality, and native language settings, P6 eventually found the right options but needed multiple explanations.

P6 seemed to understand the instructions of the log out task, and didn't ask as many questions. However, there was a brief moment of confusion before realizing the meaning of "log out" as "stop."

Participant 7

P7 managed to find the "create account" option but initially entered only 4 digits instead of the required 6 characters.

During the conversation with Emely, P7 appeared happy and engaged. P7 started the conversation promptly but needed assistance with using the microphone and sending messages. As the conversation progressed, P7 became more comfortable. However, P7 faced difficulties in understanding and asking questions.

P7 acknowledged understanding the microphone and send button but didn't grasp the concept of clicking other elements on the page. Further, P7 found it challenging to understand the purpose of the suggestion button, particularly the representation of a lightning. P7 eventually gave up and couldn't understand how to find the progress statistics. P7 also struggled to comprehend the purpose of the buttons and didn't realize they were interactive elements.
Finally, P7 encountered difficulties and didn't understand how to log out.

Participant 8

P8 appeared to be slightly uncomfortable throughout the session. When creating an account, P8 seemed hesitant and encountered difficulties, requiring assistance to enter the correct email format.

During the conversation with Emely, P8 continued to feel uneasy and mentioned that it was the first time using Emely on the phone. P8 faced challenges in understanding Emely's fast speech and struggled to respond appropriately. There were moments where P8 misunderstood Emely's questions, and provided unrelated answers. Despite our instructions, P8 had difficulty taking the initiative to ask own questions and move the conversation forward.

5.2.3. Interviews

Question 1

P6 and P7 both mentioned that it was their first time using Emely. P8, on the other hand, had used Emely twice before, but not at home. Only at AMF and on a tablet.

Question 2

When asked about their preference between talking and writing with Emely, all participants preferred talking, as they found it difficult to write or read.

Question 3

P6 expressed that using Emely was not difficult and they found it enjoyable. They described Emely as kind and mentioned feeling happy while using it. P7 also found it enjoyable but mentioned that it could be a bit challenging to speak.

Question 4

Regarding the ease or difficulty of certain tasks, P6 found logging out and accessing settings to be hard. P7 did not provide a response to this question.

Question 5

Suggestions for improving Emely to make it easier to use were not specified by either participant.

Question 6

P7 mentioned that the best thing about Emely was the ability to talk and receive responses. Regarding Emely's avatar, P7 described it as kind and pleasant. P6 also had positive views about Emely's avatar, mentioning that it looked happy and kind. P6 expressed a desire to use Emely at home, every day, particularly after school, to have someone to talk to and not feel lonely. They both stated that Emely was very good. When asked if Emely should look different, P6 answered negatively, stating that Emely's current appearance was great, being happy and kind.

5.2.4. Questionnaires

SUS

Unfortunately, the System Usability Scale (SUS) could not be accurately measured in the second test session. The participants' understanding of the tasks and their ability to provide meaningful responses caused doubts about the validity and reliability of the SUS questionnaire results. It appeared that all participants struggled to grasp the purpose and requirements of each task, which might have influenced their responses to the SUS items. Due to this limitation, we cannot confidently use the result obtained from this session.

NPS

The measurement of the Net Promoter Score (NPS) proved challenging due to the participants' evident lack of understanding when responding to the questionnaire. It became apparent that the participants did not fully comprehend the meaning or implications of their answers, which could have significantly affected the NPS results. Due to this limitation, the result obtained from this session cannot confidently be used.

5.3. Analysis

In the account creation task, two out of three were able to successfully complete it with some guidance, while one participant encountered difficulties. Looking at these specific

observations, it can discern broader implications. The difficulties faced by participants during account creation point to potential usability issues or areas that could be improved in Emely's user interface. The struggles and confusion experienced by participants suggest that the interface or instructions may not be clear enough. To enhance the user experience, simplifying the account creation process, providing clearer instructions, or improving the user interface design could be beneficial. Additionally, these findings indicate the importance of offering additional user support or guidance during the onboarding process.

During the conversation with Emely, all three participants faced challenges in understanding and formulating questions to varying degrees. Additionally, one of the participants decided to discontinue the tasks after the conversation with Emely. Taking a broader perspective, the results of the conversation task highlight certain implications. Participants' difficulties in understanding instructions, asking questions, and responding appropriately indicate potential areas for improvement in Emely's conversational interface. Enhancing the clarity of instructions, providing additional guidance for question formulation, and improving Emely's ability to understand varied accents could enhance the user experience. Moreover, the varying levels of comfort and engagement observed among participants indicate the importance of providing a user-friendly and intuitive interface. Clearer instructions on microphone usage and message sending could help users navigate these features more easily, increasing their confidence and involvement in the conversation.

In the task of finding buttons and settings, two out of the three participants performed the required actions. Among them, one participant struggled to comprehend the purpose of buttons, while the other participant did not provide specific information about their performance in this task. These observations point to potential usability issues in the button and settings interface of Emely. It is crucial to improve the clarity and intuitiveness of button representations, especially when they may not have direct visual associations with their intended functions. Providing the buttons clearer could help users better understand their purpose and functionality.

Regarding the task of logging out, two out of the three participants attempted to complete it. One participant initially experienced confusion but eventually understood the instructions, while the other participant encountered difficulties in comprehending how to log out. The analysis of the participants' performance in the logging out task presents a challenge in determining the specific factor that contributed to their difficulties. It is unclear whether the issues encountered by the participants can be attributed to shortcomings in the interface design or a lack of communication. One of them experienced a brief moment of confusion before comprehending the meaning of "log out" as "stop." The successful completion of the task upon indicating the word "stop" suggests that the difficulty stemmed from a communication problem rather than any inherent flaw in the interface. Therefore, the ambiguous nature of the participants' struggles necessitates caution in drawing definitive conclusions regarding the underlying cause.

The overall validity and reliability of the results have been subject to questioning, primarily due to concerns regarding the participants' understanding of the tasks. It remains uncertain whether the participants truly grasped the purpose and requirements of each task or merely claimed to understand without a clear understanding of the assigned objectives. The findings suggest that the task with the highest success rate was the conversation with Emely. This outcome could potentially be attributed to the participants engaging in direct communication with Emely, as verbal interaction may have facilitated a greater level of task comprehension. This limitation raises questions regarding the extent to which the outcomes were influenced by participants' reading limitations and their ability to fully engage with the tasks as intended.

No correlation can be drawn between prior usage of Emely and performance in the test. It was evident in test round 2 that the participant (P8) who had previous experience with Emely performed the poorest. Interestingly, P8 also had the lowest language proficiency among the participants in the second test round. This suggests that prior experience with Emely did not necessarily lead to better results in the test. Other factors, such as language proficiency, may have had a greater impact on the participant's performance.

6. Discussion

The research problem addressed in this study centers around the crucial role of language proficiency, specifically Swedish, in acquiring employment opportunities. Emely, the conversational agent under investigation, aims to facilitate second language acquisition and improve conversational skills, which are vital for individuals seeking employment. By providing a platform for practicing dialogues and learning Swedish, Emely aims to support the target group in their transition to the working world. The key research questions focus on the design implications for enhancing the user experience and usability of Emely, identifying the challenges and limitations in its current user interface, and proposing improvements to enhance efficiency and satisfaction for the intended target group.

In the first test round, participants generally found Emely to be a helpful conversational agent for language learning, regardless of their preferred mode of communication. Emely's conversational intelligence and contextual retention capabilities appeared to contribute to the positive user experience reported by the participants. The importance of adapting the conversation to the user type and providing multiple communication options to accommodate individual preferences and needs was highlighted. However, language barriers, particularly during account creation and registration, were identified as potential usability challenges. Additionally, certain tasks, such as finding progress statistics and voice settings, proved challenging, indicating a need for improvement in Emely's graphical user interface.

During test session two, a notable observation was the significance of understanding in communication. The test group had significantly lower proficiency in the Swedish language than the first test group. This finding raises concerns regarding the validity of the test session, consequently questioning its reliability. A major key finding from test session two is the inadequacy of the interface for individuals with limited reading abilities. The fact that at least three out of the total eight participants, counting both first and second test round, mentioned their inability to read suggests the need for interface improvements and adaptations to cater to this user group as well.

The findings suggest how important it is to design Emely and similar tools in a way that accommodates individuals who are not able to read. In order to make language learning accessible to a wider range of users, it is crucial to ensure that those who cannot read are also able to use and practice a second language effectively. By including features and functionalities that accommodate different learning abilities, such as audio instructions, visual cues, and interactive elements, it can create a more inclusive and equitable language learning experience. This inclusive design approach will not only benefit individuals with reading difficulties or disabilities but also contribute to a more comprehensive and effective learning environment for all users.

The challenges in communication between researchers and the target group due to language barriers have most certainly affected the validity of the results. While efforts were made to anticipate potential scenarios and provide clarifications, it is difficult to account for every possible misunderstanding. The absence of an interpreter necessitated the use of translation tools and clarification from the moderator, but this approach has not completely mitigated the risk of misunderstandings. Additionally, the focus on the primary objective of improving the user experience within the given time constraints limited the in-depth analysis of individual aspects.

Based on the findings of the study, future research could be designed for users with limited reading abilities. Considering the significant number of participants who mentioned their inability to read, it is crucial to improve the interface to accommodate users who rely on alternative modes of communication. This may involve incorporating visual cues, audio prompts, or other non-textual elements to enhance usability for this target group. Simplify the testing process for improved data collection. To ensure the collection of more reliable data, it is recommended to simplify the testing procedures and adapt them for the target group.

Future research could specifically examine the influence of conversation on the user experience of conversational agents like Emely. By analyzing the dynamics and quality of the conversations, researchers can gain insights into how different conversational aspects affect user satisfaction, engagement, and language learning outcomes. This would provide valuable information for further improving the design and functionality of conversational agents. By implementing these recommendations, it is possible to enhance the usability and user experience of Emely, making it more effective and accessible for language learners.

7. Conclusion

This study focused on addressing the research problem of language proficiency, particularly in Swedish, and its impact on employment opportunities. The conversational agent Emely was examined as a tool for facilitating second language acquisition and improving conversational skills, which are crucial for individuals seeking employment. Through the analysis of participant feedback and observations during test sessions, valuable insights were gained regarding the design implications, challenges, and limitations of Emely's user interface. The findings emphasized the need for interface improvements, accommodating users with limited reading abilities, and considering language barriers and proficiency levels in the design process. Despite the limitations imposed by language barriers and time constraints, the study provided valuable insights for enhancing the usability and user experience of Emely and similar tools.

7.1. Future directions

Building upon the findings and limitations of this study, there are several avenues for future research and development. First, future research should focus on improving the interface to cater to users with limited reading abilities. This involves exploring alternative communication modes, such as visual cues and audio prompts, to enhance usability for this user group. Additionally, simplifying the testing process and adapting it for the target group can lead to more reliable data collection and insights. Researchers should also investigate the influence of conversation on the user experience of conversational agents like Emely. Analyzing conversation dynamics and quality can provide valuable insights into user satisfaction, engagement, and language learning outcomes. This information can guide further improvements in the design and functionality of conversational agents. Finally, future research should aim to enhance the usability, accessibility, and effectiveness of Emely, ultimately supporting language learners in their journey towards improved language proficiency and increased employment opportunities.

8. References

[1] SCB, “Olika förutsättningar på arbetsmarknaden för inrikes och utrikes födda”, 2023. [Online]. Available: https://www.scb.se/hitta-statistik/statistik-efter-amne/arbetsmarknad/arbetskraftsundersokningar/arbetskraftsundersokningarna-aku/pong/statistiknyhet/arbetskraftsundersokningarna-aku-2023/ (accessed Apr. 17, 2023).

[2] S. Ruan et al., “EnglishBot: An AI-Powered Conversational System for Second Language Learning” *26th International Conference on Intelligent User Interfaces*, Apr. 2021, doi: https://doi.org/10.1145/3397481.3450648.

[3] M. Altinkaya and A. W. M. Smeulders, “Assisted Speech to Enable Second Language” *Proceedings of the 1st International Workshop on Multimodal Conversational AI*, Oct. 2020, doi: https://doi.org/10.1145/3423325.3423735.

[4] W. Huang, K. F. Hew, and L. K. Fryer, “Chatbots for Language Learning—Are they really useful? A systematic review of chatbot‐supported language learning” *Journal of Computer Assisted Learning*, vol. 38, no. 1, Feb. 2022, doi: https://doi.org/10.1111/jcal.12610.

[5] G. C. Guerino and N. M. C. Valentim, “Usability and User eXperience Evaluation of Conversational Systems: A Systematic Mapping Study”, *Proceedings of the 34th Brazilian Symposium on Software Engineering*, Oct. 2020, doi: https://doi.org/10.1145/3422392.3422421.

[6] NordAxon, “Creation of MLaaS NLP-driven language assistant”. [Online]. Available: https://www.nordaxon.com/cases/helsingborg-stad-creation-of-mlaas-nlp-driven-language-assistant (accessed Apr. 17, 2023).

[7] SCB, “Arbetsmarknaden för utrikes födda i Sverige och EU - AKU tilläggsundersökning 2021”, 2023. [Online]. Available: https://www.scb.se/contentassets/7c51b8e58aad44fa81a12fbd772d1db0/am0401_2023a01_amftbr2301.pdf (accessed May. 18, 2023).

[8] IBM, “What is Conversational AI”. [Online]. Available: https://www.ibm.com/topics/conversational-ai (accessed Apr.17, 2023).

[9] M. Mctear, Conversational AI : Dialogue Systems, Conversational Agents, and Chatbots. San Rafael, Ca: Morgan & Claypool Publishers, 2021.

[10] S. Singh and H. Beniwal, “A survey on near-human conversational agents” *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 10, Nov. 2021, doi: https://doi.org/10.1016/j.jksuci.2021.10.013.

[11] M. M. Mariani, N. Hashemi, and J. Wirtz, “Artificial intelligence empowered conversational agents: A systematic literature review and research agenda” *Journal of Business Research*, vol. 161, Jun. 2023, doi: https://doi.org/10.1016/j.jbusres.2023.113838.

[12] M. A. Kuhail, N. Alturki, S. Alramlawi, and K. Alhejori, “Interacting with educational chatbots: A systematic review” *Education and Information Technologies*, no. 28, pp. 973–1018, Jul. 2022, doi: https://doi.org/10.1007/s10639-022-11177-3.

[13] K. Seaborn, N. P. Miyake, P. Pennefather, and M. Otake-Matsuura, “Voice in Human–Agent Interaction,” *ACM Computing Surveys*, vol. 54, no. 4, pp. 1–43, May 2021, doi: https://doi.org/10.1145/3386867.

[14] H. D. Brown, *Principles of language learning and teaching*, 6th ed. Upper Saddle River: Pearson, 2014.

[15] ISO - International Organization for Standardization, “*ISO 9241-210:2019 - Ergonomics of human-system interaction: Human-centered design for interactive systems*”, 2019. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:9241:-210:ed-2:v1:en (accessed Apr. 17, 2023).

[16] T. Tullis and W. Albert, *Measuring the user experience : collecting, analyzing, and presenting usability metrics*. Amsterdam: Morgan Kaufmann, 2013.

[17] ISO - International Organization for Standardization, “*ISO 9241-11:2018 - Ergonomics of human-system interaction — Part 11: Usability: Definitions and concepts*”, 2018. [Online]. Available: https://www.iso.org/standard/63500.html

[18] K. Moran, “Usability Testing 101,” *Nielsen Norman Group*. [Online]. Dec. 2019. Available: https://www.nngroup.com/articles/usability-testing-101/ (accessed Apr. 18, 2023).

[19] J. Nielsen, “10 Heuristics for User Interface Design,” *Nielsen Norman Group*. [Online]. Nov. 2020. https://www.nngroup.com/articles/ten-usability-heuristics/ (accessed Apr. 18, 2023).

[20] E. Hall, *Conversational design*. New York, NY, USA: A Book Apart, 2018.

[21] A. Følstad and M. Skjuve, “Chatbots for customer service: user experience and motivation” *Proceedings of the 1st International Conference on Conversational User Interfaces - CUI ’19*, 2019, doi: https://doi.org/10.1145/3342775.3342784.

[22] Q. Vera Liao, W. Geyer, M. Muller, and Yasaman Khazaen, “Conversational Interfaces for Information Search,” *Understanding and improving information search.*, pp. 267–287, Jan. 2020, doi: https://doi.org/10.1007/978-3-030-38825-6_13.

[23] E. Ruane, S. Farrell, and A. Ventresque, “User Perception of Text-Based Chatbot Personality”, *Chatbot Research and Design*, pp. 32–47, 2021, doi: https://doi.org/10.1007/978-3-030-68288-0_3.

[24] A. Murgia, D. Janssens, S. Demeyer, and B. Vasilescu, “Among the Machines: Human-Bot Interaction on Social Q&A Websites”, *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA ’16*, pp. 1272–1279, 2016, doi: https://doi.org/10.1145/2851581.2892311.

[25] I. Medhi Thies, N. Menon, S. Magapu, M. Subramony, and J. O’Neill, “How Do You Want Your Chatbot? An Exploratory Wizard-of-Oz Study with Young, Urban Indians” *Human-Computer Interaction - INTERACT 2017*, pp. 441–459, 2017, doi: https://doi.org/10.1007/978-3-319-67744-6_28.

[26] T. L. Smestad and F. Volden, “Chatbot Personalities Matters,” *Internet Science*, pp. 170–181, 2019, doi: https://doi.org/10.1007/978-3-030-17705-8_15.

[27] E. J. de Visser et al., “Almost human: Anthropomorphism increases trust resilience in cognitive agents” *Journal of Experimental Psychology: Applied*, vol. 22, no. 3, pp. 331–349, Sep. 2016, doi: https://doi.org/10.1037/xap0000092.

[28] L. Clark et al., “What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents” *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI ’19*, no. 475, pp. 1–12, May 2019, doi: https://doi.org/10.1145/3290605.3300705.

[29] C. Murad and C. Munteanu, “Designing Voice Interfaces: Back to the (Curriculum) Basics,” *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, Apr. 2020, doi: https://doi.org/10.1145/3313831.3376522.

[30] E. Adamopoulou and L. Moussiades, “Chatbots: History, technology, and applications,” *Machine Learning with Applications*, vol. 2, Dec. 2020, doi: https://doi.org/10.1016/j.mlwa.2020.100006.

[31] W. Galvan-Romero, “The Role of Intelligent Personal Assistants in Migrant Learners' Willingness to Communicate in English as a Second Language,” *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, Apr. 2022, doi: https://doi.org/10.1145/3491101.3503818.

[32] B. J. Oates, *Researching information systems and computing*. London, UK, SAGE, 2006.

[33] J. Lazar, J. H. Feng, and H. Hochheiser, *Research Methods in Human-Computer Interaction*, 2nd ed., Cambridge, MA, USA: Elsevier Science & Technology, 2017. [Online]. Available: https://eds.p.ebscohost.com/eds/detail/detail?vid=2&sid=18ef634a-545e-46a3-895f-791e8

73ebd55%40redis&bdata=JkF1dGhUeXBlPWlwLGNvb2tpZSx1cmwsc2hpYiZsYW5nPX
N2JnNpdGU9ZWRzLWxpdmUmc2NvcGU9c2l0ZQ%3d%3d#AN=malmo.b2410973&db=c
at05074a (accessed Apr. 18, 2023)

[34] D. Ghosh, P. S. Foong, S. Zhang, and S. Zhao, "Assessing the Utility of the System Usability Scale for Evaluating Voice-based User Interfaces," *Proceedings of the Sixth International Symposium of Chinese CHI*, pp. 11–15, Apr. 2018, doi: https://doi.org/10.1145/3202667.3204844.

[35] N. I. Fisher and R. E. Kordupleski, "Good and bad market research: A critical review of Net Promoter Score," *Applied Stochastic Models in Business and Industry*, vol. 35, no. 1, pp. 138–151, Nov. 2018, doi: https://doi.org/10.1002/asmb.2417.

[36] J. Nielsen, *Usability Engineering*. Amsterdam: Morgan Kaufmann, 1994.

[37] J. Nielsen, "Thinking Aloud: The #1 Usability Tool," *Nielsen Norman Group*. [Online]. Jan. 2012. Available: https://www.nngroup.com/articles/thinking-aloud-the-1-usability-tool/ (accessed Apr. 18, 2023).

[38] M. J. Wenger and J. H. Spyridakis, "The relevance of reliability and validity to usability testing" *IEEE Transactions on Professional Communication*, vol. 32, no. 4, pp. 265–271, 1989, doi: https://doi.org/10.1109/47.44538.

[39] J. Nielsen, "10 Usability Heuristics for User Interface Design". [Online]. *Nielsen Norman Group*. [Online]. Nov. 2020. Available: https://www.nngroup.com/articles/ten-usability-heuristics/ (accessed May. 18, 2023).

Appendix

Appendix A

Test Script - för testpersonen (in Swedish)

Instruktioner (Läses upp för testpersonen)

Hej! Det är jag som är Jonna, och jag som är Julia. Vi studerar på Malmö Universitet och skriver vårt examensarbete hos NordAxon som har gjort Emely. Vi är här för att få feedback så att Emely kan bli ännu bättre. Vi kommer tillsammans med er kolla på de olika sakerna som man kan göra i Emely.

Vi vill bara vara tydliga med att vi inte testar er, vi testar produkten och designen. Om något är konstigt eller svårt, är det något vi behöver veta så att vi kan göra Emely enklare att använda. Dela gärna med dig av allt ni tänker på – både bra och dåligt. All din feedback kommer att vara till hjälp för oss.

Vi vill också vara tydliga med att alla svar som samlas in under testet kommer att behandlas anonymt och all information som samlas in kommer att raderas i slutet av studien.

För att kunna få ut så mycket och bra information som möjligt hade vi gärna velat spela in skärmen samt ljud för att kunna gå igenom det efter testet. Är det okej?

Testet

- Ni kommer att få gå igenom aktiviteter och ni kommer få information om en åt gången.
- Ni ska, i den mån det går, försöka utföra uppgifterna själva. Skulle uppgiften kännas helt omöjlig att lösa, be om hjälp. Ni får be testledaren upprepa uppgiften och instruktionerna vid behov, men fråga gärna innan varje uppgift om det är något som är otydligt.
- När du anser att du klarat din uppgift så informerar du instruktören om att du är färdig. Till exempel genom att säga “klar”.
- När testet är slutfört kommer vi att ställa några snabba frågor och så ska du få fylla i en enkät och svara på några frågor. Svaren kommer användas för att försöka förbättra upplevelsen för Emelys användare.

Uppgifter

1. Skapa konto
2. Ha en konversation med Emely (indelad i 3 uppgifter)
3. Hitta hjälpknappen
4. Hitta förslagsknappen
5. Hitta till statistik av framsteg
6. Hitta inställningar för rösthastighet, ljudkvalitet och modersmål
7. Logga ut

Test Script - for moderator/observator (in Swedish)

Instruktioner

- Varje uppgift ska tas tid på
    - Uppgiften börjar när testledaren läst upp varje enskild uppgift högt
    - Uppgiften avslutas när testpersonen säger att de har klarat uppgiften

- Instruktioner ska läsas högljutt för användaren
    - Allmän information innan testet påbörjas
    - Innan varje ny, vara tydlig med att testpersonen förstår instruktionerna

Uppgifter

1. Skapa konto
    - **Success**
        i. Användaren säger att den skapat konto och loggas in
    - **Success med vägledning**
        i. Användaren klarar uppgiften med vägledning från testledaren
    - **Fail**
        i. Testpersonen säger att hen inte kan skapa ett konto
        ii. Instruktören ser att testpersonen är fel ute och uppenbart förvirrad
        iii. Testpersonen tror att de är klara med uppgiften men instruktören bedömer att testpersonen fyllt i informationen på ett felaktigt sätt

2. Ha en konversation med Emely
    - Notera vilken av konversationerna testpersonerna väljer.
    - Notera om testpersonen använder förstaspråk eller andraspråk.
    - Notera om testpersonen skriver eller pratar med Emely.
    - Notera om konversationen är i rätt kontext; svarar testpersonen på det Emely säger och vice versa.
    - Notera hur testpersonen påverkas om kontexten blir fel.

    A. Välja konversation
    - **Success**
        i. Testpersonen säger att hen startar en av tre valbara konversationer med
        Emely
    - **Success med vägledning**
        i. Användaren klarar uppgiften med vägledning från testledaren
    - **Fail**
        i. Testpersonen säger att hen inte vet hur man påbörjar en konversation
        ii. Instruktören ser att testpersonen är fel ute och uppenbart förvirrad
        iii. Det tar längre tid än 1 min att slutföra uppgiften (alltså välja)

B. Påbörja en konversation med Emely

- **Success**
  - **i.** Testpersonen svarar på det Emely säger minst 2 ggr, antingen med text eller tal
- **Success med vägledning**
  i. Användaren klarar uppgiften med vägledning från testledaren
- **Fail**
  - i. Testpersonen svarar endast 1 gång
  - ii. Instruktören ser att testpersonen är fel ute och uppenbart förvirrad
  - iii. Testpersonen misslyckas med att svara på det Emely säger

C. Slutföra konversationen med Emely

- **Success**
  - **i.** Testpersonen slutför en hel konversation med Emely; den räknas som slut när Emely säger hejdå.
- **Success med vägledning**
  - i. Användaren klarar uppgiften med vägledning från testledaren
- **Fail**
  - i. Testpersonen lämnar konversationen utan att slutföra den
  - ii. Instruktören ser att testpersonen är fel ute och uppenbart förvirrad
  - iii. Testpersonen tror att de är klara med uppgiften men instruktören bedömer att testpersonen utfört uppgiften på ett felaktigt sätt

3. Hitta hjälp-knappen (konversation)
   Innan uppgiften börjar, be testpersonen gå in på valfri konversation Nämn för testpersonen att det finns möjlighet att få hjälp med funktionerna här och be dem hitta knappen.

   - **Success**
     - **i.** Testpersonen säger att hen hittat knappen (med frågetecken på)
   - **Sucess med vägledning**
     - i. Användaren klarar uppgiften med vägledning från testledaren
   - **Fail**
     - i. Testpersonen säger att de inte kan hitta hjälp-knappen
     - ii. Instruktören ser att testpersonen är fel ute och uppenbart förvirrad
     - iii. Testpersonen tror att de är klara med uppgiften men instruktören bedömer att testpersonen utfört uppgiften på ett felaktigt sätt

4. Hitta förslagsknappen (konversation)
   Innan uppgiften börjar, be testpersonen gå in på valfri konversation Nämn för testpersonen att det finns möjlighet att få förslag på vad hen kan svara Emely och be dem hitta knappen.

   - **Success**
     - **i.** Testpersonen säger att hen hittat knappen (med glödlampan på)
   - **Success med vägledning**
     - i. Användaren klarar uppgiften med vägledning från testledaren
   - **Fail**
     - i. Testpersonen säger att de inte kan hitta förslags-knappen
     - ii. Instruktören ser att testpersonen är fel ute och uppenbart förvirrad
     - iii. Testpersonen tror att de är klara med uppgiften men instruktören bedömer att testpersonen utfört uppgiften på ett felaktigt sätt

5. Hitta statistik över framsteg
   Om testpersonen inte är på startsidan, be hen gå dit. Förklara att framsteg, såsom antalet ord och nya ord för varje konversation sparas så att hen ska kunna följa sin utveckling. Be testpersonen hitta till "Framsteg" under "Mitt konto" i Menyn.

   - **Success**
     - **i.** Testpersonen säger att hen hittat statistiken (som finns under meny - mitt konto - framsteg
   - **Success med vägledning**
     - i. Användaren klarar uppgiften med vägledning från testledaren
   - **Fail**
     - i. Testpersonen säger att de inte kan hitta statistiken över sina framsteg
     - ii. Instruktören ser att testpersonen är fel ute och uppenbart förvirrad
     - iii. Det tar längre tid än 3 min att hitta uppgiften
     - iv. Testpersonen tror att de är klara med uppgiften men instruktören bedömer att testpersonen utfört uppgiften/hittat informationen på ett felaktigt sätt

6. Hitta inställningar för rösthastighet, ljudkvalitet och modersmål
   Om testpersonen inte är på startsidan, be hen gå dit. Förklara att det finns inställningar för rösthastighet, ljudkvalitet och modersmål om testpersonen skulle vilja ändra dessa. Be testpersonen hitta "Inställningar" i menyn.

   - **Success**
     - **i.** Testpersonen säger att hen hittat knappen (med kugghjulet på)
   - **Success med vägledning**
     - i. Användaren klarar uppgiften med vägledning från testledaren

- **Fail**
    i. Testpersonen säger att de inte kan hitta hjälp-knappen
    ii. Instruktören ser att testpersonen är fel ute och uppenbart förvirrad
    iii. Det tar längre tid än 3 min att hitta uppgiften
    iv. Testpersonen tror att de är klara med uppgiften men instruktören bedömer att testpersonen utfört uppgiften på ett felaktigt sätt

Appendix B

SUS questionnaire (in Swedish)

1. Jag tror att jag skulle vilja använda Emely regelbundet *

	1	2	3	4	5	
Håller inte alls med	○	○	○	○	○	Håller helt med

2. Jag tycker att Emely är mer komplicerad än den behöver vara *

	1	2	3	4	5	
Håller inte alls med	○	○	○	○	○	Håller helt med

3. Jag tycker Emely är lätt att använda *

	1	2	3	4	5	
Håller inte alls med	○	○	○	○	○	Håller helt med

4. Jag tror att jag skulle behöva personlig support för att kunna använda Emely *

	1	2	3	4	5	
Håller inte alls med	○	○	○	○	○	Håller helt med

5. Jag tycker att Emely's funktioner fungerar bra *

	1	2	3	4	5	
Håller inte alls med	○	○	○	○	○	Håller helt med

6. Jag tycker att det finns många saker som inte är logiska med Emely *

	1	2	3	4	5	
Håller inte alls med	○	○	○	○	○	Håller helt med

7. Jag tror att de flesta skulle kunna lära sig att använda Emely ganska snabbt *

	1	2	3	4	5	
Håller inte alls med	○	○	○	○	○	Håller helt med

8. Jag tycker att Emely är besvärlig att använda *

	1	2	3	4	5	
Håller inte alls med	○	○	○	○	○	Håller helt med

9. Jag känner mig väldigt säker och trygg (på vad jag gör) när jag använder Emely *

	1	2	3	4	5	
Håller inte alls med	○	○	○	○	○	Håller helt med

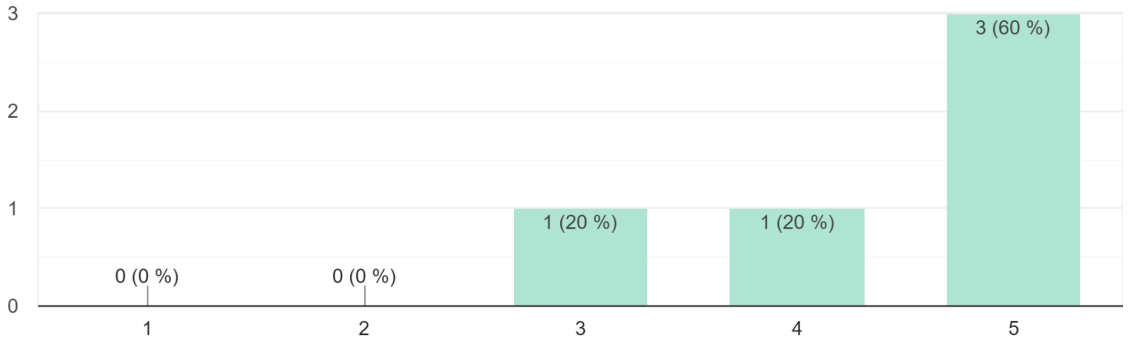
10. Jag behöver lära mig ganska mycket innan jag kan börja använda Emely *

	1	2	3	4	5	
Håller inte alls med	○	○	○	○	○	Håller helt med

SUS questionnaire answers - First test round (in Swedish)

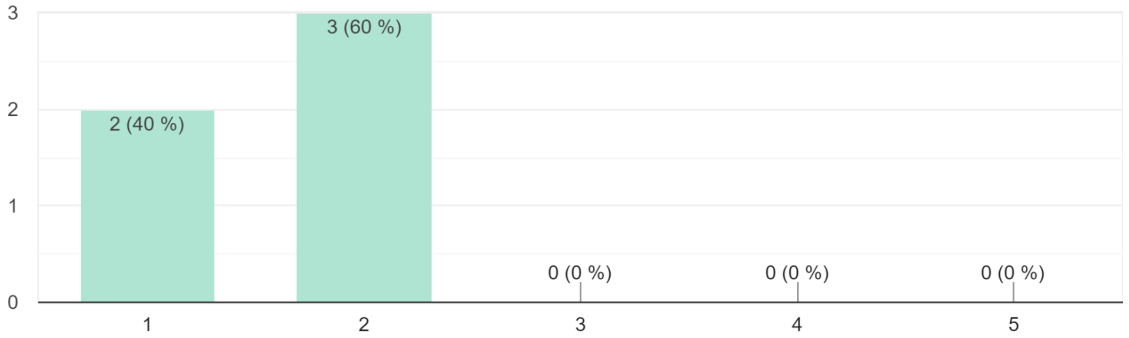
1. Jag tror att jag skulle vilja använda Emely regelbundet

5 svar



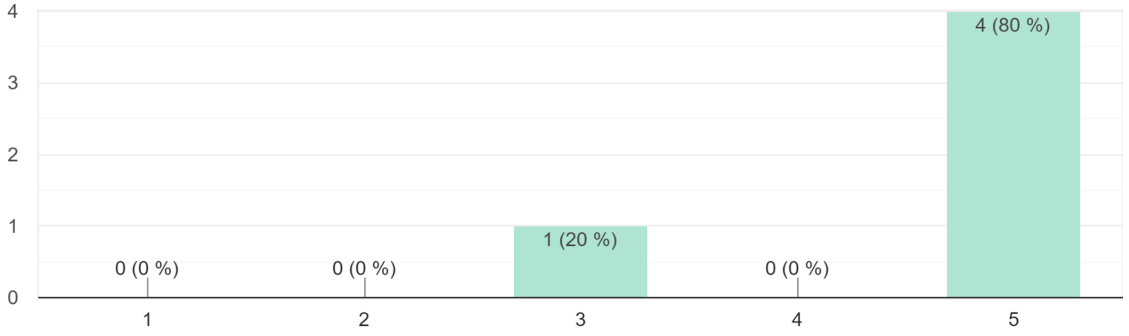
2. Jag tycker att Emely är mer komplicerad än den behöver vara

5 svar



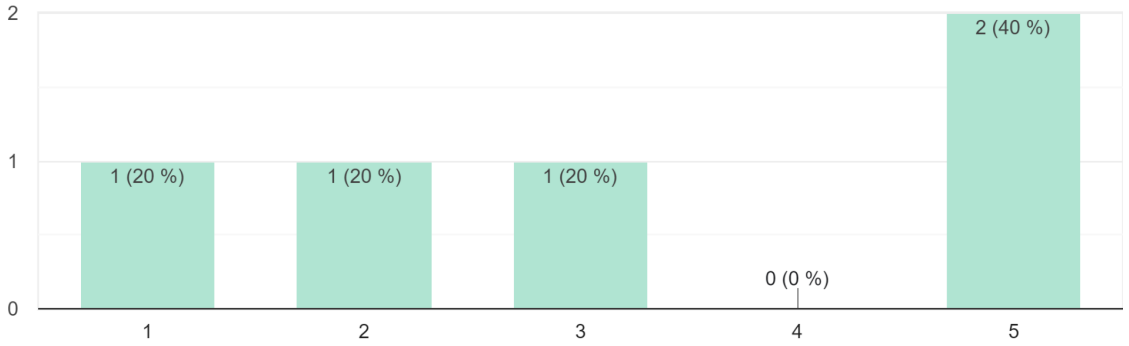
3. Jag tycker Emely är lätt att använda

5 svar



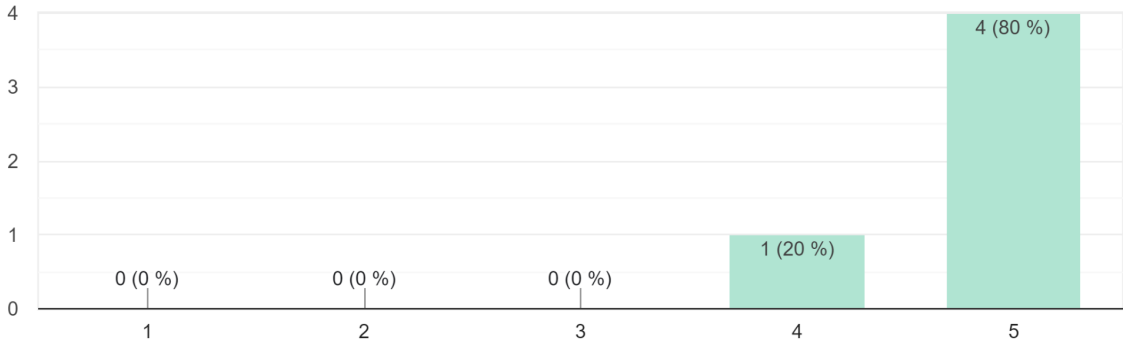
4. Jag tror att jag skulle behöva personlig support för att kunna använda Emely

5 svar



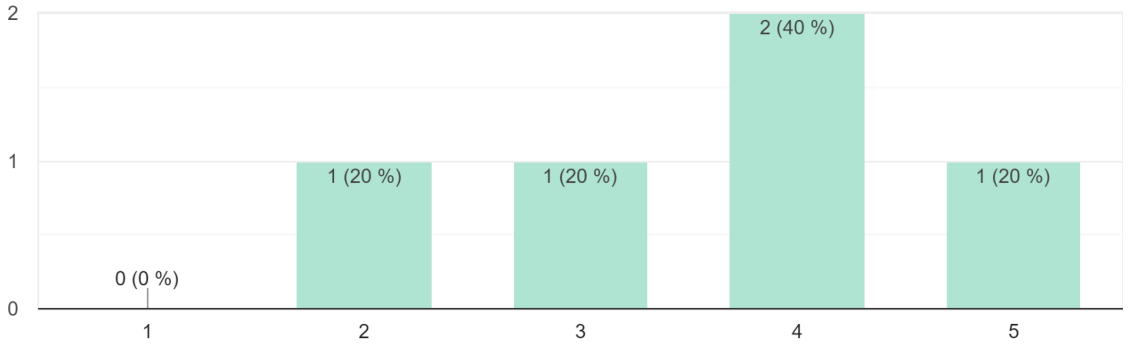
5. Jag tycker att Emely's funktioner fungerar bra

5 svar



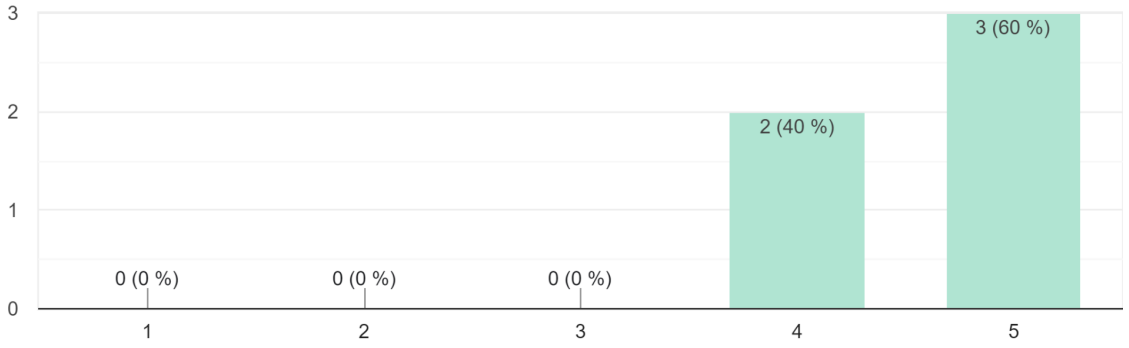
6. Jag tycker att det finns många saker som inte är logiska med Emely

5 svar



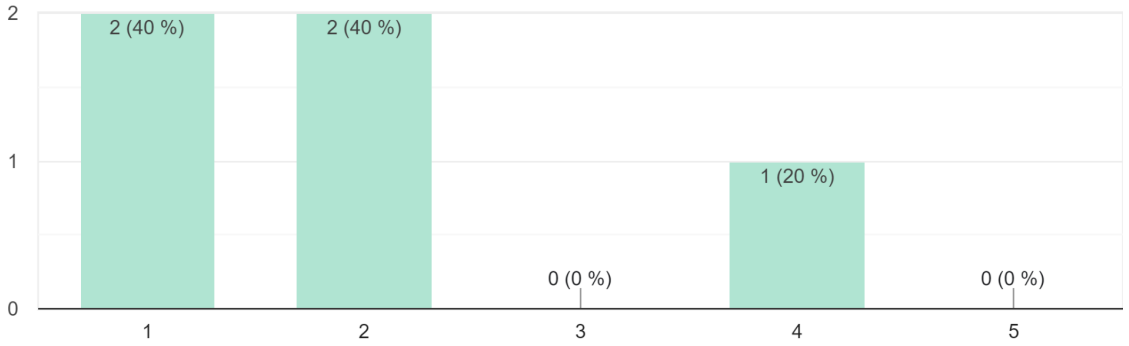
7. Jag tror att de flesta skulle kunna lära sig att använda Emely ganska snabbt

5 svar



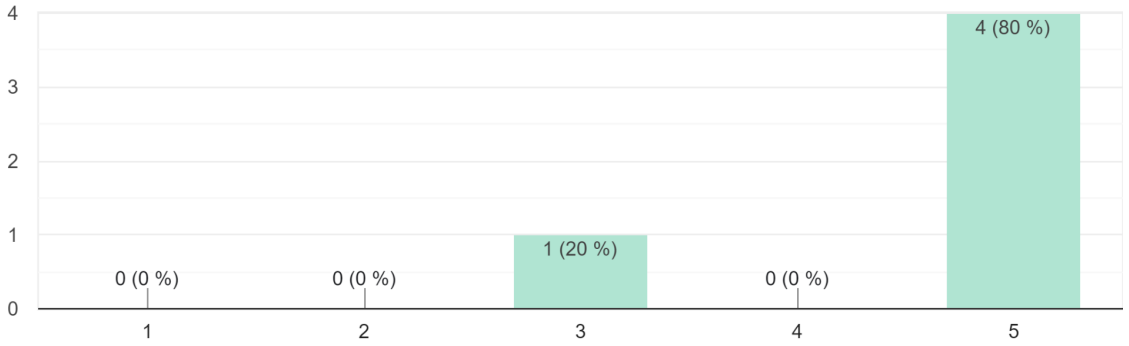
8. Jag tycker att Emely är besvärlig att använda

5 svar



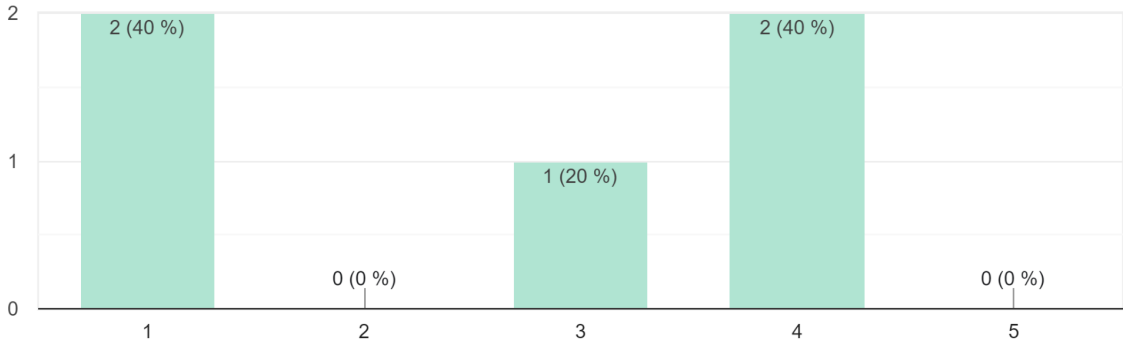
9. Jag känner mig väldigt säker och trygg (på vad jag gör) när jag använder Emely

5 svar

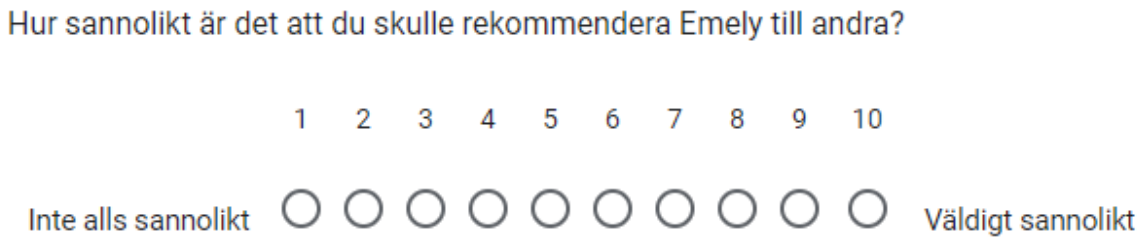


10. Jag behöver lära mig ganska mycket innan jag kan börja använda Emely

5 svar



NPS question (in Swedish)



NPS question answers (in Swedish)

Hur sannolikt är det att du skulle rekommendera Emely till andra?

5 svar