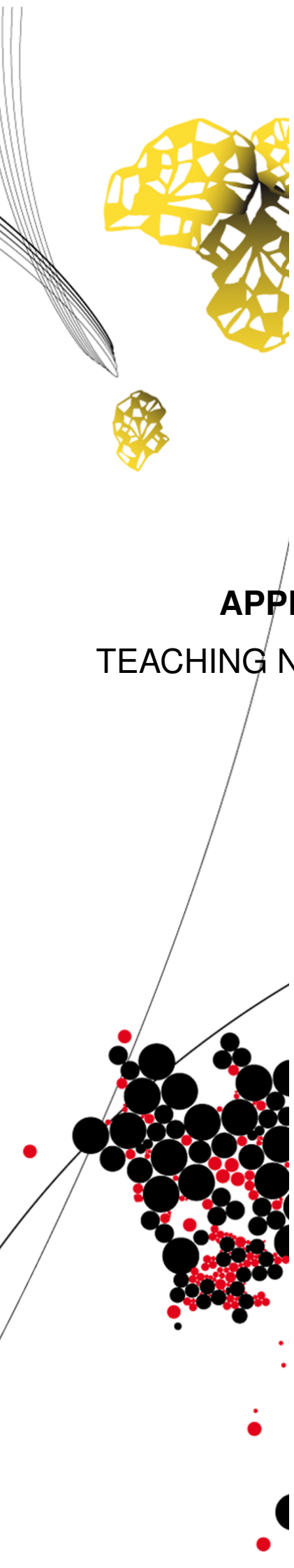


UNIVERSITY OF TWENTE.

Faculty of Electrical Engineering, Mathematics & Computer Science

APPLYING E-LEARNING AND PERSUASIVE DESIGN:

TEACHING NEW USERS OF AN ONLINE ACCOUNTING TOOL THE BASICS OF ONLINE BOOKKEEPING

T.A. BRUS

MSc Thesis, December 2017

Supervisors:
Dr. M. Theune (University of Twente)
Dr. R. Klaasen (University of Twente)

External:
S. van Bruijnsvoort-Kamies, MSc (Moneybird)
J.C.G Weeink, BSc (Moneybird)

Telecommunication Engineering Group
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente
P.O. Box 217
7500 AE Enschede
The Netherlands

This master thesis originally contained some confidential data about the company at which the research was conducted (Moneybird). These data have been removed (parts of Tables 5.7 and 5.8). A version of this thesis that does contain these data can be requested via the supervisors of this project, as listed on the front page.

Abstract

A key aspect of running a successful software company, is not only crafting a well-designed product, but even more so educating its users on how to use the software effectively. In this research, we focus on Moneybird, a web-based bookkeeping application for entrepreneurs with over 150.000 users. We determine that there is great potential for entrepreneurs to improve their bookkeeping skills, and that this directly affects the perceived value of the product to the end user, and in turn the long-term success of the company. A user that is more proficient in bookkeeping experiences greater productivity, is less likely to make mistakes, and is more likely to recommend the software to a friend.

This research study addresses the use of e-learning and persuasive technology to design an online bookkeeping training for users of Moneybird. We focus on newly signed up users of Moneybird since this group likely has little bookkeeping experience, and thus the greatest potential to learn. We study e-learning, a research area that focuses on evaluating the effectiveness of various ways in which digital learning material can be presented, and ultimately extracting these findings into numerous design principles that are straightforward to apply in practice. Furthermore, we discuss persuasive technology: the use of technology to nudge (*persuade*) users in performing a certain behaviour, preferably a behaviour of which they are already convinced is beneficial to them. An existing framework for persuasive design is discussed and a number of design principles from this framework are directly applied in the current research. Based on the theoretical foundation, an online learning environment is developed that incorporates the evidence-based design principles of both research fields. In four interactive lessons, the basics of online bookkeeping are explained. The lessons consist of simulations of bookkeeping activities, that show cause and effect, similar to the actions that can be performed in Moneybird itself. Furthermore, the lessons contain multiple choice questions, that test whether the users understood the learning material that was presented, and provide feedback when incorrect answers are given. Each lesson ends with an overview of the key points of the lessons. To evaluate the effectiveness of the learning environment, two experiments are conducted: a usability test and an A/B test.

In the first experiment, the usability test, six people new to Moneybird were invited to partake in sessions of one hour, during which the test facilitator observed the behaviour of these users as they navigated through the learning environment. The goal of the experiment was to spot usability issues that could then be fixed before the application would be made available to a larger audience, in the final experiment. An overview of problems that were encountered was constructed, and appropriate solutions were discussed and implemented. Overall, the findings of the usability test were positive, with a few minor issues that were encountered, and no critical issues that would require the help of the test facilitator for users to continue, were found. In the second experiment, the A/B test, users were randomly split into two groups. Using the existing event tracking architecture of Moneybird people that signed up to Moneybird were tracked anonymously. The test group was given access to the learning environment while the control group was not. We hypothesized that users in the test group would be more active in the main application as a result of a better understanding of bookkeeping, gained through the learning environment.

Contrary to our hypothesis, the collected data showed that, across the various metrics, users were slightly less active in the main application when provided access to the learning environment. We hypothesize that this is because new users have a fixed amount of time they are willing to invest, and since the test group had access to two applications (the main application and the learning environment) they spent some of their time in the learning environment and as a result less hereof in the main application. The thesis ends with a number of recommendations for future work, emphasizing the value of qualitative research through surveys and user interviews.

Acknowledgements

I would like to thank my supervisors, Mariët Theune and Randy Klaassen, for their support and valuable feedback during my graduation period. I would also like to thank my supervisors at Moneybird for the many ideas and discussions held, and for the opportunity to conduct and complete my final thesis at their office.

Contents

1. Introduction

In her well-received book [SB15], Kathy Sierra explains the benefits for companies to *make users awesome*: a focus on increasing proficiency of users with the product (as opposed to simply making the product better), which results in greater engagement and increased brand loyalty. An example of this is free photographing tutorials that often come with buying a camera, with the goal to turn consumers into better photographers. This in turn results in consumers that more easily buy upgrades, such as expensive lenses, which are needed to progress as a photographer. In this thesis, we review existing research studies in order to bring this idea into practice at a company that offers bookkeeping software: Moneybird.

Moneybird is a company founded in 2008 with the mission to help entrepreneurs manage their bookkeeping in an easy and fun way. This is accomplished via web-based software that has evolved over time and now supports all the aspects of a complete bookkeeping. Currently, about 150.000 entrepreneurs have signed up for an account at Moneybird. Paid subscription plans are available that remove limitations on the number of invoices and incoming documents users can add per month. Aside from that, all functionality is available to all users. In short, entrepreneurs use the product in the following way:

- Invoices and estimates are created and sent to the client. The application keeps track of their status, the generated revenue, and taxes.
- Incoming documents, such as receipts and sales invoices, are submitted and categorized. Moneybird automatically keeps track of the charges and VAT.
- Bank transactions are processed and categorized. Moneybird automatically updates the balance.
- Additional bookings, such as payroll entries or a starting balance, are added through the *memorial*[1].
- Reports, such as the income statement and balance sheet, are automatically generated and provide detailed insight into the financial health of the company. These reports also form the basis for tax declarations and annual returns.

The users of Moneybird vary in how much they know about bookkeeping and this may influence their attitude towards the product. This presents an opportunity for the company to follow up on the suggestions in Sierra's book and to increase the bookkeeping proficiency of its users. This could be done via an online training. Offering an online learning platform which increases the bookkeeping proficiency of users with Moneybird is expected to have the following benefits:

1. A potential decrease in the number of questions received at the support desk.
2. Entrepreneurs that understand how a decent bookkeeping works are more dependent on the software and will stay longer as a customer.

[1] A manual debit or credit entry in the journal of ones bookkeeping [Mon17].

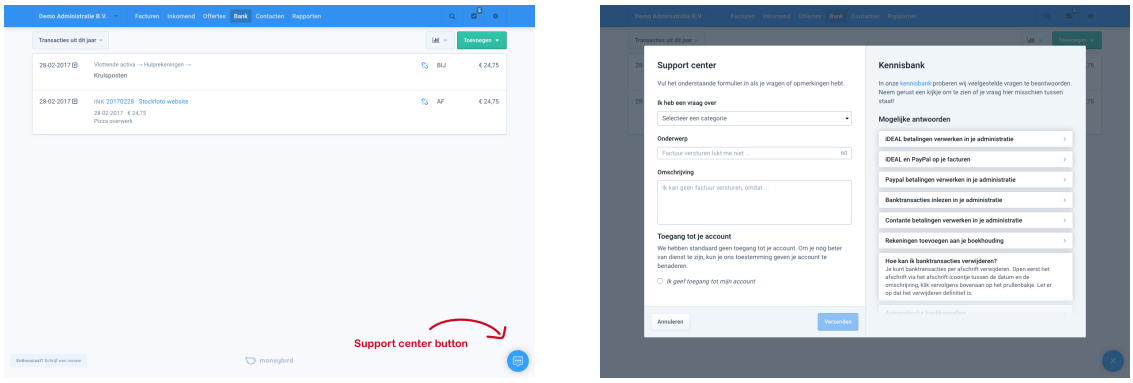
3. A more positive attitude towards the product is developed because users experience less frustration and gain greater satisfaction from getting their work done in a correct and efficient manner.

A quick search reveals that a large volume of paid and freemium bookkeeping training is already available in a broad range of formats: via books, videos and live training. It is clear that a market exists for this kind of material. However, from the company's experience, the observation is that still only few entrepreneurs make use of these options (be it for economical reasons, lack of motivation, or lack of time), even though it could benefit them. In this research, this issue is counteracted by designing a bookkeeping training that incorporates evidence-based principles of online learning, and by tailoring the training to the needs of the users of Moneybird. In particular the research fields persuasive technology and gamification can help motivate users to partake in an online training. An additional benefit we can provide is tailoring the training to the target audience of freelancers and small business owners, and to create a bridge between the training and the product. Incorporating evidence-based principles of online learning is essential for designing an effective learning environment. The goal of this research is to explore existing work related to online learning experiences, and to bring this into practice in the form of an online bookkeeping training, and finally to evaluate its effectiveness with real end users.

In the following sections we further establish the context in which the research takes place, by means of an analysis of the existing tools via which users can currently learn about bookkeeping in Moneybird, and by means of an analysis of the target audience and their learning objectives.

1.1. Existing learning tools at Moneybird

There are a number of ways in which users can learn more about bookkeeping in Moneybird. When first using the application, users are shown an onboarding screen that contains a welcoming message as well as a number of steps they can take to customize their administration. Furthermore, over the course of the next two weeks, a series of emails with helpful tips are sent to the user to further explain the most important topics, such as processing received invoices and keeping track of bank transactions. Aside from this, users are free to explore the rest of the application. Since most of the features of the product, such as creating an invoice or adding a new contact, have a limited effect and can be easily reverted, the user can relatively safely explore the application and its functionality.



(a) Support center button

(b) Support center

Figure 1.1.: From any page in the application, users can open up the support center. Here they can either directly ask a specific question or explore relevant knowledge base articles. The articles are selected based on the current page and the question category selected by the user.

Finally, if the user has questions, they can do three things. Firstly they can visit the knowledge base[2] which contains articles on all features of Moneybird. The articles often start with some general information about the topic and bookkeeping and ends with an explanation how the given described problem is solved in the application, often accompanied with an explanatory screenshot. The second way via which users can get help is the support center. The support center can be reached from any page via the blue button, as shown in Figure 1.1. Relevant knowledge base articles are shown as well as a form for users to directly ask their question to the support team. Questions may also be asked via social media or email. Finally, the financial advisors program[3] offered by Moneybird allows users to directly ask their bookkeeping question to their advisor if they have any.

1.2. Target audience and learning objectives

Within the Netherlands, every person with entrepreneurial intentions is a potential user of Moneybird, and is thus potentially helped by an online training about bookkeeping. From a practical point of view it makes sense in this research to focus on the people that have already decided to use Moneybird as their bookkeeping tool, or those that have just signed up, and are thus easy to reach. Currently over 150.000 people have signed up for an account. It is hard to precisely describe the user group because it is so diverse, but in general it can be said that the users are entrepreneurs: small business owners, freelancers, contractors, and so on. From the company's experience, users are generally fairly well educated but usually they lack specific education regarding bookkeeping. Aside from these regular users, a second group of users exists: financial advisors. Financial advisors are part of the aforementioned advisors program that allows users to find a bookkeeping expert in order to get some help with their administration. Currently about 40 advisors are part of the program, and can be found on the website. The advisor can help users by periodically verifying that everything is in order, or by aiding in any other way a regular accountant does. Financial advisors are screened before they can join the program and always have a background in bookkeeping or accounting, and possess a high level of financial literacy. They are experts in their field. Finally, there are a number of ways in which all users differ, which influences the design of an online bookkeeping training:

[2] https://www.moneybird.nl/ontdek/kennisbank/
[3] https://www.moneybird.nl/adviseurs/

1. Financial literacy: *what is their current level of bookkeeping knowledge?*
2. Engagement & adherence: *how often do they use Moneybird, and are they still active?*
3. Moneybird proficiency: *are they familiar with the interface and are they able to locate its various functionality?*

Because of the ways in which users can differ, it makes sense for this research to choose a specific target audience, preferably one that is easy to differentiate and and still has a clear potential to learn. This has led to the decision to focus on the target audience of new users. The existing architecture of the application allows for easily tracking this group of users, and furthermore, these users have not been previously influenced by any experience with the application, so in that respect they are a blank slate. They may have different levels of prior bookkeeping knowledge, though. This issue is mostly addressed by the concept of "learner control" and specifically by allowing learners to choose their own pace [CM16]. Furthermore, if the focus would be on experienced users, it would be hard to tell what their learning objectives are since some users may not know everything, but that can still be sufficient for what they are trying to accomplish. So, what this means is that for new users there will almost always be a potential to learn something new, albeit just about the interface and not even about bookkeeping, while for experienced users the potential to learn something new (related to bookkeeping or the interface) is much smaller.

Finally, it is difficult to provide a detailed view of the learning objectives of the chosen target audience. To do that a large scale survey would need to be conducted first. However, we do know what all the different things are a user can do in the main application, and we know, with the help of domain experts, what bookkeeping knowledge is required to achieve proficiency at those bookkeeping tasks. Still, the learning objectives may be different for each individual, and only through further research it will become clear how ambitious each individuals' goals are. Nevertheless, an argument can be made for aiming to make users as proficient at bookkeeping as possible, which is in essence what this research is really about.

1.3. Research questions

As will become clear in the next chapter, the goal of this research study is to make a contribution to the field of e-learning and persuasive technology by applying the relevant literature in a new context: web-based bookkeeping software for entrepreneurs. By implementing and evaluating the effectiveness of an online training, we intend to eventually draw conclusions that are relevant to both fields of research. Furthermore, it seems that the application of e-learning and persuasive technology to teach entrepreneurs (as opposed to high school or university students) about bookkeeping is unique among existing literature. This leads to the following research questions that are to be answered in this study:

1. How can e-learning and persuasive system design techniques be employed to build an engaging online learning experience that increases the proficiency of users in doing bookkeeping via Moneybird?
    a) What conclusions can we draw on the effect of the online training in relation to bookkeeping proficiency?
    b) What conclusions can we draw on the effect of (the presence of) the online training in relation to user engagement?

1.4. Research outline

Let us provide a short overview of this thesis. In the following chapter, Chapter 2, we lay the theoretical foundation for the rest of this research study by reviewing two research fields: *e-learning*, and *persuasive technology*. Relevant work in the form of existing online learning environments is reviewed as well. In Chapter 3, we discuss the design phase, in which we explore the numerous considerations and decisions made while designing an online learning experience targeted at our goal of increasing the bookkeeping proficiency of users of Moneybird. We motivate our design choices by the literature and the related work that was discussed in the previous chapter. To further improve the learning experience, we conduct a small scale usability test (Chapter 4). In Chapter 5, we evaluate the learning environment among newly signed up users of Moneybird in an online experiment (A/B test). During this experiment, anonymous data is collected that reflects the behaviour of the end users in the learning environment and Moneybird itself. A number of measurable hypotheses are formulated to answer the research questions. In the final part of this research, we analyze and interpret the data collected during the experiment. We show which hypotheses are rejected and which are not. Furthermore, we use the collected data to provide an overview of the usage characteristics of the learning environment. Finally, in Chapter 6, we summarize and discuss the results of this research study, and provide a short discussion on the methodology used, as well as some recommendations for future work.

2. Background

With this research we intend to design an online training based on proven techniques and principles. In this section we explore several research fields that are relevant to our goal of designing an online learning experience. Human computer interaction is the overlapping research field that first comes to mind, but also cognitive psychology and pedagogy are relevant research fields. The combination of these fields is sometimes referred to as *the science of instruction*, and when applied in the digital world this is known as *e-learning*. Furthermore, in order to design an engaging online learning experience, we review literature in the field of *persuasive technology*. Persuasive technology is based around the idea that the thoughts and behaviors of people can be influenced through interaction with technology. A specific approach to persuasive technology is *gamification*. All of these concepts and more are discussed in the following sections. Finally, we conclude the chapter with an overview of related work: other existing examples of online learning environments.

2.1. Literature review

When designing educational content for the web, one notices that a lot of opportunities appear for creating more appealing and engaging content, such as via animations, videos, and interactive exercises (as opposed to what is possible via traditional printed media). But what do we know about designing educational content for digital media? In the following section we discuss the existing literature on *e-learning*: "training delivered on a digital device such as a smart phone or a laptop computer that is designed to support individual learning or organizational performance goals" [CM16]. In relation to the current research, on an individual level, the goal is for users to gain a better understanding of bookkeeping. On an organizational level, the goals are to ultimately reduce the number of support questions, to increase engagement, and to convert more users to customers.

2.1.1. E-learning

The first research area that we discuss is that of e-learning. Researchers such as Mayer, Sweller, and Moreno have established a set of design principles that stimulate effective e-learning. These principles, which are discussed below, have been field tested in many different settings, for example by Harskamp et al. [HMS07], Chang and Yang [CY10], and Isa et al. [IMS+13]. In most studies conducted by Mayer and colleagues, the user group consists of high school or university students with little prior knowledge regarding the topic explained[1]. The lesson material is usually about a technical topic, such as math or physics, and is relatively short. In other studies, these principles have also shown to be effective when non-technical topics are taught [MLS95], and when taught to learners of other ages [GPM+03]. Furthermore, Mayer argues that the science of instruction is in large part defined by "the focus on authentic learning situations, rather than contrived learning situations" and he names a "computer-based training program to teach employees how to use a new database program" as an example of such an authentication situation. In conclusion, the field of e-learning is not bounded by the typical (aforementioned) educational topics.

[1] Sometimes when learners possess greater prior knowledge, contradicting results are found. This phenomenon is referred to as the *expertise effect* [SWVGVM11] [KZSK14] [BKS10] [KCS00].

Typical research on instructional techniques takes place by having a control group that learns via a standard training, and a treatment group that learns with a modified version of the training, namely extended by the given instructional technique. The effectiveness is tested via a *transfer test*, a test that evaluates the understanding of the learner with regard to the provided material, by asking questions that require the participant to apply the knowledge in a new situation. When the participant is able to do this, this is commonly referred to as "deeper learning" to have occurred [WM08]. The results of the transfer test are expressed via the *effect size*[2]. Effect sizes greater than 0.5 indicate a practical significance that make the technique worthy of considering for implementation. Sometimes a *retention test* is used as well, but this test merely asserts whether the participants remembered the content the way it was presented.

The established e-learning principles that we quickly mentioned originate from the cognitive load theory, a theory which refers to the mental effort used in the *working memory* when processing information. The theory models a person's *working memory* as a cognitive system that has limited capacity and is responsibly for temporarily holding information. Situations in which cognitive load increases are when a lot of information is presented at once, when distractions are present, or due to factors such as age, for example with small children who have less general knowledge to draw from than adults. Three different types of cognitive processing are distinguished: essential, extraneous, and generative cognitive load. The first type refers to additional cognitive processing imposed by the way information is presented. The second type refers to the inherent complexity associated to the specific topic itself. And, finally, the third type refers to the effort required to fully understand and permanently store the information that is presented. Different e-learning principles address different types of cognitive processing. For example, extraneous processing can be reduced by applying the redundancy principle, and generative processing can be stimulated via the personalization principle. In the following section we discuss these principles one by one and summarize their effectiveness across various different studies.

Principles and strategies

The main principle established in e-learning, by researchers such as Mayer, Sweller, and Moreno, is the *multimedia principle*: the assertion that people learn more deeply (that is, people performed better on transfer tests), from words and graphics than from words alone. This principle forms the basis for using multimedia instructions to foster learning. Examples of multimedia instructions are spoken and printed text, illustrations, charts, animations, videos, and simulations. Over the years, a growing consensus has risen around the multimedia principle [DFH07] [PBB+07]. To effectively apply this principle, careful thought still needs to be put in on how to design educational graphics, and how to deliver e-learning material in general. To this end, a set of evidence-based follow-up principles have been established which are discussed below. The effect sizes of these principles across various different studies are summarized in Table 2.1.

Modality principle Substantial empirical evidence has been found for the assertion that deeper learning occurs when explanatory information alongside graphics is presented as audio instead of on-screen text [Gin05] [HMS07] [MM98] [MM99]. The psychological explanation is that via this method information is split across two separate cognitive channels (auditory and visual), that can work simultaneously, which prevents a channel from being overloaded. The principle was found to be less effective when learners are increasingly familiar with the content, and when they are not native speakers of the language spoken in the audio narration [CM16]. Furthermore,

[2] Effect size is the number of standard deviations in difference between the mean scores of two groups in an experiment [Coh77].

it may not always be feasible to use audio because of the time investment required, because it is more difficult to keep up-to-date, and because it may not be suitable in the context of use of the end user.

Segmenting principle This principle states that deeper learning occurs when content is presented in logical chunks with breaks, as opposed to continuously without breaks [MC01] [May05b] [CM16]. The effectiveness of this principle depends on the complexity of the material. Furthermore, in some cases it is not possible to have a clear separation between different topics because the content is too intertwined.

Pretraining principle The pretraining principle states that for low-knowledge learners it can be beneficial to explain key concepts or terminology (*jargon*) before presenting the main content of a course. This allows beginners to more easily comprehend the material and not get overloaded by a surplus of information. Evidence for this effect was shown in a multitude of studies [MMW02] [KKvM06] [PCS02] [Ayr06] [CAS05]. The most important boundary condition here is yet again that learners are not so familiar with the domain yet.

Coherence principle Time and time again, research shows that "less is more" when it comes to adding details that are not directly relevant to the main topic of the material and thus make it less coherent [PMSB11] [MSA+14] [CM16]. These details are often referred to as *seductive details* because on the surface they seem interesting, but in reality they distract from the main content. This is especially problematic for low knowledge learners. This counter-intuitive effect of omitting content to increase learning has been shown for adding extraneous sound [MM00], pictures [But06] [MHL01], and words [HM98] [LSMH07] [MGJR08] [MHL01]. Deciding what content is essential and what content is better left out is still an open problem. However, to give an example from the aforementioned research, it is apparent that adding sound effects or background music is almost always a bad idea. Finally, the effect of the coherence principle is less pronounced when learners posses greater prior knowledge. In that case some motivational graphics may actually increase the engagement and learning effectiveness [PMSB11] [MSA+14].

Redundancy principle For native speakers, the addition of identical on-screen text ("subtitles") when using audio to explain graphics, has a negative effect on learning transfer [MM02] [SSG+14] [CS91] [CM16]. In this case the on-screen text is *redundant* to the narrated graphic. The psychological explanation here is that you avoid overloading the visual channel which is already used to process the graphic shown on the screen. The effect is most apparent when the content is fast-paced and the learner is familiar with the language of the spoken text. For learners that are not familiar with the language, adding on-screen text may actually be beneficial. Additionally, directing the learner's attention to specific parts of the graphic by using words to label key elements can also result in increased learning (this is called *signaling* [May05a] [CM16]).

Contiguity principle: spatial & temporal When using on-screen text to explain graphics, sometimes the screen size or the layout prevents a user from seeing both at the same time, resulting in them having to switch back and forth between the text and the graphic. The *spatial contiguity principle* argues that related pieces of information should be kept together as close as possible [HHH09] [May89] [MA92] [MSBM95]. An example violation of this principle is when the feedback after answering a multiple choice quiz is given on a new separate screen. Another example is when references are used to refer to key elements of a graphic, and the legend is placed somewhere

Principle	**Median Effect Size**	Studies with effect size > 0.5
Multimedia	1.50	9 of 9
Modality	0.97	20 of 21
Segmenting	0.98	3 of 3
Pretraining	0.92	7 of 7
Coherence	1.32	10 of 11
Redundancy	0.69	8 of 10
Contiguity	1.11	8 of 8
Personalization	1.30	10 of 10

Table 2.1.: Summary of the effectiveness of the various e-learning principles (taken from [CM16]). As mentioned before, techniques with an effect size less than 0.5 are usually not worth implementing.

else, far below the graphic and perhaps even outside of the screen. This same principles applies to timing and is called the *temporal contiguity principle* [MA91] [MS94] [MMBV99]. Here the recommendation is that when providing audio to narrate graphics such as a video or animation, the audio should be played simultaneously with the graphic. This seems obvious, however practitioners may be tempted to separate audio and video because that would allow for users to select the modality they prefer: visual or auditory. Clark and Mayer argue that the extraneous processing induced by having to hold all the information in memory, when watching both the graphic and the audio in sequence, offsets the supposed benefits [CM16].

Personalization principle The personalization principle states that deeper learning occurs when the learner experiences a stronger social presence, through conversational style or via the presence of a virtual coach [CM16] [Wik17a]. One might argue that this detracts from the seriousness of the lesson, however, cognitive research on discourse processing shows that people put more effort in understanding material when they (feel like) they are in a conversation, as opposed to when simply presented with information [BMS+96]. Thus, when it comes to writing style, an informal conversational style instead of a formal style is preferred [MM00] [Kar10]. Something as seemingly insignificant as changing "the" to "your" can already make a difference in learning transfer [MFFC04]. Furthermore, empirical evidence was found for using a polite tone of voice (*"you may want to try doing this thing first"*) to stimulate deeper learning for less experienced learners [WJM+08] [MDM11]. Finally, conversational agents, such as virtual coaches, may aid the learning process by further reinforcing social presence and specifically when used to point out key elements or visually demonstrate important concepts [MDM03] [MMSL01] [Atk02] [MD12] [MRO10].

This wraps up our overview of evidence-based e-learning principles. Next we will look at one additional literature study that reviews evaluation techniques used in a web-based learning setting, similar to that of the current research.

Evaluation strategies

How are e-learning systems typically evaluated? We have talked briefly about transfer tests, but are there also ways to measure skills and not only knowledge gain? And finally, what

about the learners' attitude? This subjective aspect is worth evaluating as well. Learners' attitude and their motivation is an important aspect of learning [RWT05]. In this section we look at a literature review study conducted by Chumley-Jones et al. that summarizes the evaluation strategies used across a selection of 76 *web-based learning* (WBL) papers in the medical, dental and nursing domain [CJDA02]. 41 of those papers were descriptive and 35 were evaluative. Of those 35 evaluative papers the following different evaluative aspects were identified: knowledge gain, learner attitude, learning efficiency, and program cost. We discuss the first three aspects since the one study that looked at program costs is fairly outdated (1995) and made a few assumptions that do not necessarily hold anymore today.

Knowledge gain Regarding this first aspect, twenty studies were identified that evaluate knowledge gains. The majority of these studies used a multiple-choice quiz to measure the difference in knowledge before and after the training. In two cases a more elaborate test was set up that was evaluated by an expert. An example of such a test is a "standardized patient evaluation". From the web-based learning interventions, only a couple used a control group to compare the difference in knowledge gain between web-based and traditional training material. The contents of the training were kept identical. The result was that no significant difference in knowledge gain was found between the material presented on the web and the material presented via the traditional way.

Learner attitude The next aspect that was identified across the selected papers is learner attitudes. A distinction was made between attitudes towards web-specific and content-specific attributes. Regarding web-specific attributes, it was found that properties such as accessibility, navigational structure, and attractiveness are all predictors of higher learner satisfaction. Page load time, however, was found to be the most important factor and can be as important as the content itself, according to one study at least. Another attribute that was reviewed was the use of asynchronous communication, for example by asking or receiving feedback in an asynchronous way. One study found that learners felt more comfortable asking questions this way, however in general there is insufficient evidence to tell when asynchronous communication is beneficial and when not.

When it comes to evaluating the attitude of learners towards content-specific features, some studies found high ratings for WBL material, however, when comparing the ratings with other modalities (such as traditional physical learning material), high ratings were reported here as well in several studies. In relation to the satisfaction of learners only the pedagogical soundness (whether the content was well-designed from an educational point of view) was found to be correlated. Numerous other variables such as self-rated competence with technology, and learners' age were found to not be of influence to the reported learner satisfaction. Finally, the study reports that strong evidence is found for the observation that well-designed WBL interventions increases learners' confidence.

Learning efficiency Finally, regarding learning efficiency, in one WBL intervention it was found that equivalent test scores can be achieved using web-based methods (as opposed to the typical text-based material) but in a shorter timespan. However, in a second study the effect was much less pronounced, so in this case regarding learning efficiency the conclusion is that further research is required.

As pointed out, this literature study only reviewed WBL solutions discussed in medical, dental, and nursing literature. Nevertheless, some interesting conclusions are drawn,

notably that learners generally have a positive attitude towards WBL and that page load time is such an important factor. However, it was also concluded that web-based training generally does not lead to a greater increase in knowledge gain than traditional content, an observation that is also emphasized in the book by Clark and Mayer [CM16]. What this really means is that the difference in effectiveness is made by increasing the quality of the learning material, not merely by changing the medium. Furthermore, even if the learning material remains unchanged, web-based training can still be chosen for its other benefits, such as the ability to easily change and improve the training over time.

2.1.2. Persuasive technology

A shortcoming of of e-learning from a pedagogical point of view is that it cannot transmit emotion or engage the student in the material, a role which is traditionally fulfilled by the teacher [Mun11]. The principles laid out in section 2.1.1 only address how to design effective online educational content, but little is said about how to motivate learners: a key concept for positive learning outcomes. Only the personalization principle touches on this subject. Research shows that attitude of learners and their motivation significantly influences the amount of effort they put in regularly and persistently engaging in learning activities [Ush05]. Furthermore, in the introduction we pointed out that despite the large volume of (free) bookkeeping material available, users often neglect to make use of it. It seems that many learning environments are deserted, even if they are effective in supporting learning [MLG14]. To accommodate for all of this, another research field needs to be addressed: *persuasive technology*.

Persuasive technology, a term originally coined by Fogg in 2002 [Fog02], is interactive technology that is purposefully designed to change the attitude and/or behavior of the end user through persuasion and social influence. Fogg himself is also known for his three-factor (motivation, ability, and trigger) human behavior model [Fog09]. Fogg proposed a number of persuasive system principles which have later on been adapted and categorized by Oinas-Kukkonen & Harjumaa [OKH09]. A selection of these principles that are deemed most applicable to the current study are outlined in Table 2.2, and Table 2.3, and Table 2.4. Finally, the application of persuasive technology in web-based learning environments has been discussed in numerous studies [LZMM06] [DSM+13] [Mun11] [MLG14] [DF06].

	Primary task support
Principle	*Description*
Tailoring	Information provided by the system will be more persuasive if it is tailored to the potential needs, interests, personality, usage context, or other factors relevant to a user group.
Self-monitoring	A system that keeps track of the user's performance or status supports the user in achieving goals.
Simulation	Systems that provides simulation can persuade users by enabling them to immediately observe the link between cause and effect.

Table 2.2.: Persuasive system principles, selected from the study by Oinas-Kukkonen & Harjumaa [OKH09]. The category *primary task support* refers the main activities a user carries out in a persuasive system.

Dialogue support	
Principle	*Description*
Praise	By offering praise, a system can make users more open to persuasion.
Rewards	Systems that reward target behaviors may have great persuasive powers.
Liking	A system that is visually attractive for its users is likely to be more persuasive.
Social role	If a system adopts a social role, users will more likely use it for persuasive purposes.
Expertise	A system that is viewed as incorporating expertise will have increased powers of persuasion.

Table 2.3.: Persuasive system principles, selected from the study by Oinas-Kukkonen & Harjumaa [OKH09]. *Dialogue support* revolves around the persuasive communication between the system and the end user.

Social support	
Principle	*Description*
Social comparison	System users will have a greater motivation to perform the target behavior if they can compare their performance with the performance of others.
Competition	A system can motivate users to adopt a target attitude or behavior by leveraging human beings' natural drive to compete.

Table 2.4.: Persuasive system principles, selected from the study by Oinas-Kukkonen & Harjumaa [OKH09]. *Social support* is the category of persuasive system design that concerns the incorporation of social elements to persuade a user to perform a certain action.

A parallel concept that has emerged in more recent years is *gamification*. Gamification has broadly been defined as the "use of game-design elements and game principles in non-game contexts" [DDKN11]. Such non-game contexts include (web-based) learning environments [Mun11] [MLG14] [DF06], social networks, e-commerce, health, and marketing. Typical gamification concepts include points, rewards, leaderboards, and unlockables. These concepts are also commonly mentioned in persuasive technology literature. The idea behind gamification is that it leverages people's natural desire for socializing, curiosity, mastery, competition, achievement, status, and self-expression [Lie15]. Gamification can be seen as a specific form of persuasive technology. It supports similar types of behavior change, such as increased user engagement, but accomplishes it in a different way. Gamification techniques are heavily used by companies for commercial reasons. This is why some people criticize it and say it is merely another way for large companies to exploit customers. Within gamification, some subcategories have emerged such as *serious gaming* and *game-based learning*. These two focus more on games as an environment in which people can learn. This as opposed to the incorporation of game elements in existing

applications, which is more or less what gamification is about.

In a 2014 literature study by Hamari et al., the authors reviewed the empirical effectiveness of persuasive technology (and gamification) across 95 studies [HKP14]. About half of these studies were fully quantitative, the other half used mixed evaluation methods, or were fully qualitative. The reviewed studies took place in various different contexts, such as: health and exercise, ecological consumption, and education. Hamari et al. conclude that persuasive technologies do indeed persuade, and a positive effect was found across many studies measuring a multitude of different psychological factors. The review study ends with an interesting observation:

> — "It seems that persuasive technologies are implemented especially in contexts wherein people would be willing to undertake the target activities but find it difficult to start or continue working toward them. (...) This notion lends support to the idea that an important aspect with persuasive technologies and gamification is **whether the encouraged activity is something the user is trying to accomplish regardless** of the system or the user is instead persuaded toward a behavior that is valuable only for the designer of the system."

From this observation we learn that persuasive technology is most effective when the user is already convinced of the benefits of the target behavior.

Finally, let us discuss two problems regarding the application of persuasive technology and gamification in particular. Persuasive technology is often used to increase user engagement via techniques such as rewarding the user, offering praise, and by providing (artificial) challenges. A pitfall of these techniques is that it may distract users from their original learning objective, and that achieving these things may become goal of its own. Aside from that, Clark & Mayer point out the difference between behavioral engagement (for example: putting letters in the correct order to form an answer) and psychological engagement ("cognitive processing of content in ways that lead to acquisition of new knowledge and skills") [CM08] [CM16]. They further emphasize that high levels of behavioral activity do not necessarily lead to the desired high levels of psychological activity needed for learning. Nevertheless, by carefully applying the principles of persuasive technology around the main learning material, and by not making learning material interactive just for the sake of it, these techniques can still be applied successfully.

2.1.3. Conclusion

In the past sections we learned about the field-tested principles for e-learning that exist and are relatively straight forward to apply in practice. We learned that words and graphics as opposed to words alone lead to deeper learning. Aside from that, the general consensus is to keep things simple and to try not to overload the user mentally. If possible, different channels (visual and auditory) should be utilized, instead of presenting a lot of information via the same channel. We further learned how e-learning systems are typically evaluated: mostly via quantitative tests such as a pre- and post test consisting of multiple choice questions to assess knowledge gains. It was found that e-learning systems generally do not provide an increase in knowledge gain when compared to traditional methods, and when the content is the same. Furthermore, from a review study we learned the most important design factors that impact learner satisfaction: page load time, accessibility, navigational structure and attractiveness. Regarding content-specific features, pedagogical soundness (whether the course was well-designed from an educational point of view) was found to be the most important variable in relation to learner satisfaction, according to the review study.

We discussed persuasive technology as an outcome for motivating users to invest time in e-learning systems, and we saw that learner attitude plays a significant role in this process. Several ways to influence learner attitude through persuasive technology are shown, such as the use of rewards and praise. How to implement these in an online bookkeeping training is still an open problem, but we did state that it is most likely a good idea to keep the main content intact when implementing gamification and/or persuasive technology features. Furthermore we learned that persuasive technology seems to work best when applied in a situation where the user is already convinced of the benefits of the training, but needs some extra motivation to commit to their learning goal.

2.2. Related work

Many online learning environments have been developed, and some became very successful, serving millions of visitors every month. These platforms, such as Duolingo, Babbel, and Khan Academy, have a big impact on the way people learn on the web, since they have such a vast audience. The (design) choices that these companies made have undoubtedly contributed to their success. We can draw inspiration from these examples during the design of our own online training. Some of these companies publish research studies, something which is also useful in relation to the current research. In this section we will describe the various successful online learning platforms, and summarize the research that they conducted. Finally, we end with an example of a free online bookkeeping training already available, and discuss its strong and weak points. But first, let us start with three research studies that specifically address the design or evaluation of an (online) bookkeeping training.

2.2.1. Research studies

A few studies specifically address the implementation and/or evaluation of an (online) learning environment that teaches accounting, bookkeeping, or financial literacy. One of those studies is a study by Maynard et al. in which the authors discuss whether video games can build financial capability [MMP+12]. The case study revolves around a previously developed casual video game that intends to increase the financial capability and confidence of its users in the real world. A five-step approach was used that corresponds to the stages-of-change model (also known as transtheoretical model) of behavior change by Prochaska. The results suggest that the game is successful at engaging consumers and increasing financial literacy, which in turn could translate to better decisions in real-world situations.

In another study by Concannon et al., the authors conducted research on the attitude of Irish university students towards use of e-learning in an accounting class [CFC05]. According to the study, universities are increasingly investing in digital learning environments to further increase the quality of their education. Qualitative as well as quantitative techniques were used to gauge the experience of students with the online learning tools. The findings of the study indicate that students had a positive attitude towards the learning environment and used it as an additional resource on top of the traditional methods. The negative feedback that was received was mostly related to technical problems.

Finally, in a master thesis at the New Mexico State University, an existing experiential gaming model was used to benchmark four online financial literacy games with regards to whether they promote engagement and support active learning [Che13]. The author conducted a thorough qualitative analysis of which games accomplished these goals and which games did not. One of the observations made is that "gamification of standardized testing is what should be prevented in financial literacy games", an observation similar to our conclusion at the end of section 2.1.2. In this section we concluded that increased

behavioral engagement (such as that encouraged by gamification), can hinder the cognitive engagement needed to acquire new knowledge and skills.

2.2.2. Duolingo

Duolingo is an online platform for learning a new language. The language courses are provided for free, with some minor advertisements shown here and there. Duolingo originally started as a way to teach people a foreign language while at the same time having them translate sentences for paid purposes. However, this idea of paid translations was later on abandoned. Currently the project is backed by a number of investors and via an optional paid subscription that users can take to hide advertisements.

Duolingo works as follows. After signing up, users can select the language they wish to learn. To make progress the user must complete various lessons about different topics (see Figure 2.1). By taking a placement test, the basics can be skipped. The tests involve translating sentences from the native language to the foreign language and vice versa. The sentences are presented via on-screen text or as audio. Furthermore, the lessons involve some theory that the user can review first, after which a quiz is presented. The quiz is purposely made accessibly for people with varying skill levels. Quizzes consist of multiple choice questions that are accompanied with pictures and spoken text, which makes it even easier for users to pick the correct answer. As the user advances more difficult topics are addressed and less hand-holding is provided. Finally, a number of gamification techniques are implemented, such as the ability for users to earn experience points and *lingots*, the currency of Duolingo. This currency allows for unlocking special courses and quizzes. The experience points are displayed on the public profile of the user.

The Duolingo project was started at the Carnegie Mellon University in Pittsburgh by professor Luis von Ahn and graduate student Severin Hacker and continues to uphold a heavy science-based approach to online learning. A number of internal as well as independent studies have been conducted about the learning platform and its effectiveness. Aside from that, blog posts are frequently published with behind-the-scenes explanations of how certain features work. For example, in one article the team summarizes research they conducted that resulted in a *strength meter*: a meter that indicates to the user how skilled he or she is at a certain topic [Set16] [SM16]. This indicator decreases as time passes by, and increases each time the user practices that topic. Users can choose to redo specific exercises that they had trouble with. Aside from the numerous internal studies, an effectiveness study has been conducted by a research team at the City University of New York [VG12]. This study was not entirely independent since the funding was still provided by Duolingo. The study focused on determining the time it takes for users to learn Spanish (that is, reach the proficiency of a typical university student after a first-year beginners course, for which 130 hours are allocated). They found that it takes 34 hours to reach similar reading and writing ability. In another study by one of the same authors, it was found that it takes users 55 to 66 hours to reach the same level via Rosetta Stone (a paid online learning platform) [Ves09].

2.2.3. Babbel

Babbel is another online language learning platform and was founded in 2007. The site offers one lesson free of charge for every language, after that paid subscription plans are available. The interface of Babbel is very straightforward and has little distractions. Users can take quizzes by simply selecting the correct option or by answer questions via speech recognition. A large number of different courses are provided. For example, users can learn about grammar, vocabulary, but also about idioms and proverbs. Furthermore, courses for different dialects of a language are available, as well as courses that teach you how to

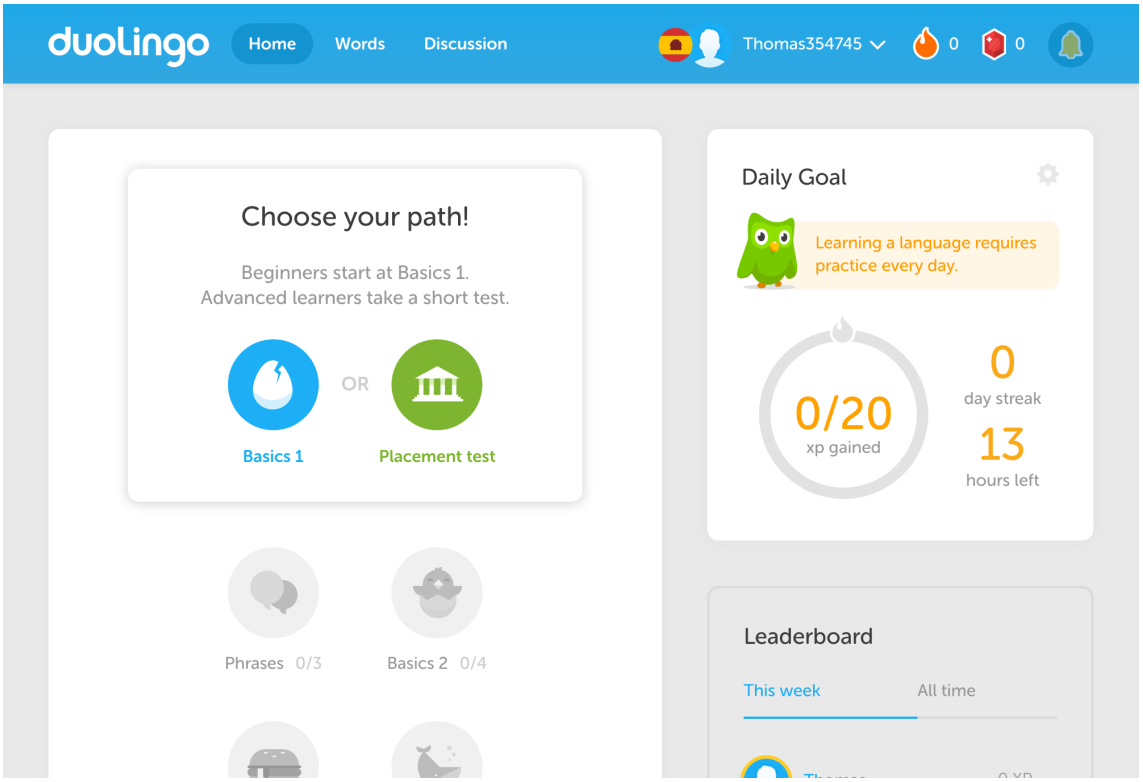


Figure 2.1.: Duolingo.

write in a formal or informal tone of voice (also see Figure 2.2). In contrast to Duolingo, Babbel has fewer features that contribute to keeping the user engaged, but does invest a lot of time and energy in putting together original learning content, as the company has over 100 language experts employed for exactly this job [Woo17]. This focus on content over engagement could be explained by the business model. Where Duolingo thrives when it has many active users, Babbel is more focused on providing their paid users with quality content. A user that has spent money on an online course might already have enough incentive to spend time on the platform and return regularly.

Not much research is available on Babbel specifically, however one study conducted by Stevenson and Liu in 2010 looked at the use of collaborative learning tools and social networking features amongst three online language learning platforms: Live Mocha, Palabea, and Babbel [SL10]. The former two have shut down by now unfortunately. The goal of the study was to investigate the role of the aforementioned "Web 2.0" technologies in relation to learning and social purposes. The methods used were usability tests and an online survey. The concept of Web 2.0 was particularly popular around the time the study was conducted, and the authors differentiate between Web 1.0 and Web 2.0 learning websites as follows:

> — "Web 1.0 users would read static content created by 'experts' who had the technical ability to write and post content [Ebn07]. Web 1.0 is contrasted with Web 2.0 in which general users consume, create, and edit content while easily collaborating with other users [ML07]. Web 2.0 tools provide users the opportunity to play a more active role of potential author, contributor, editor, or specialist."

A link exists between e-learning as we know it today and learning websites that incorporate Web 2.0 technology and this link is further explained in the aforementioned research

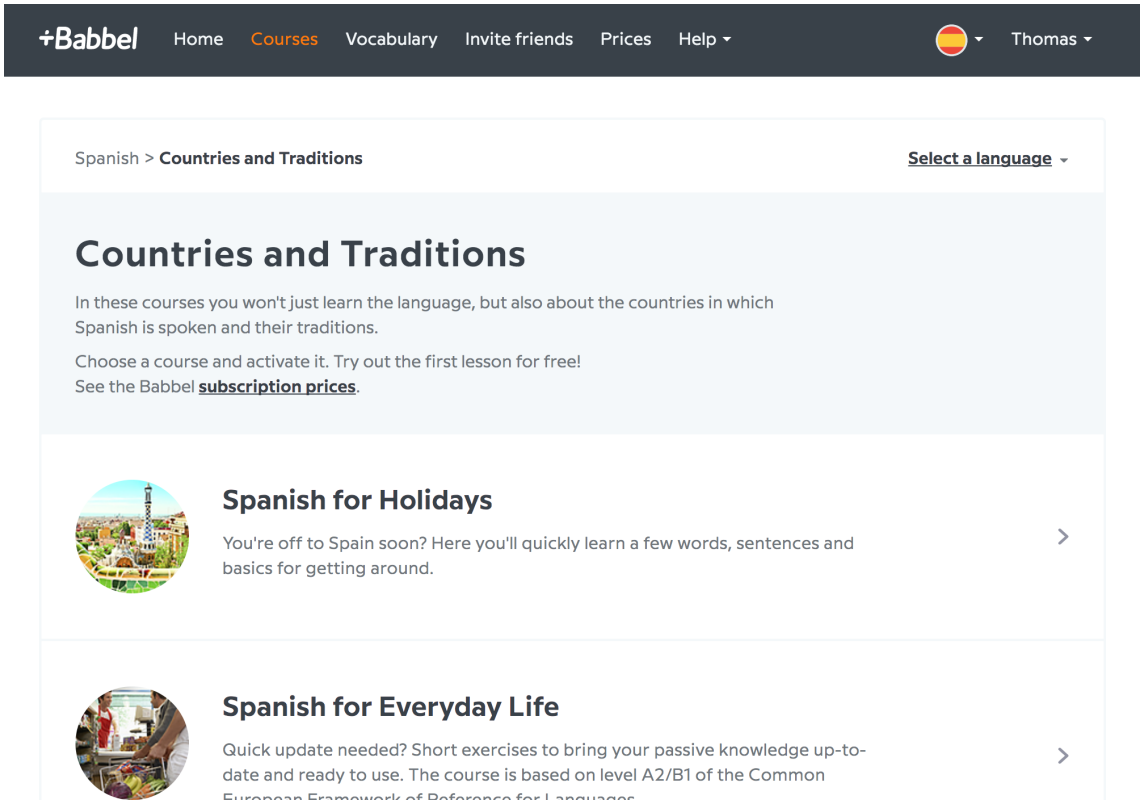


Figure 2.2.: Language course on Babbel related to Spanish traditions and countries in which the language is spoken.

article by Ebner [Ebn07]. Finally, returning to the research by Stevenson and Liu, the authors conclude the paper with a number of observations that are interesting to the current research as well:

- There is a still a place for presenting learning material using traditional (Web 1.0) techniques (see section 2.2.5 for an example), however, to be successful, learning websites may also need to provide "appealing Web 2.0 collaborative interactive tools, and a design that is convincing, interesting, and easy to use for adult users who wish to learn and collaborate with other language learners".
- The usability (technical and pedagogical) of the websites affected whether the users were likely to return and thus whether they achieve their language learning goals.
- The appearance (specifically when, according to the surveys, the site looked "young" or "cartoony") affected whether adult learners thought they could accomplish their learning goals on the site in an effective way.
- Positive feedback was received for sites that made it easy for users to accomplish their learning goals, while negative feedback was received mostly in relation to confusing or distracting interface elements that were present on the sites.

2.2.4. Khan Academy

Khan Academy is a non-profit organization founded in 2006 by Salman Khan that offers an online learning platform through their website[3]. The site has differentiated itself, amongst

[3] https://www.khanacademy.org/

other things, by making all content available for free and without the need for registration.

Content is provided on subjects such as math, science, engineering, programming, and economics. The level of the content mostly ranges from kindergarten to 8th grade high school, although recently more advanced topics have been added as well. The way Khan Academy presents its learning material is mainly via YouTube videos, which is also how it got started originally. The videos show a black canvas which the narrator uses to draw on and explain the topic at hand, as shown in Figure 2.3. This resembles the traditional way in which teachers present in front of the class room. The video can be paused and fast-forwarded, and has subtitles enabled by default. A transcript of the video is available as well, which can be used to jump to specific parts. On Khan Academy, a subject is divided into classes, which are in turn explained via multiple tutorials. A tutorial usually starts with a video explaining the concept, sometimes followed by an additional textual explanation, and often concluded by an exercise or multiple choice quiz. Finally, registered users can track their progress and can earn badges when completing exercises. To reward users for watching the introductory videos, *energy points* are awarded. If a user has trouble completing an exercise, they are encouraged to rewatch any of the videos or to reveal a hint. Revealing a hint, however, prevents the exercise from counting towards one's progress. By successfully completing an exercise, users unlock badges and avatars which are displayed on the user's public profile.

Some research has been done on Khan Academy, in particular at the Universidad Carlos III de Madrid by Muñz-Merino et al. In one study [MMVK13], the authors build a model for inferring high level parameters from low level data in an attempt to gain insight in questions such as: *can this student be motivated by gamification techniques?* They concluded that in their analysis they were unable to say whether users were motivated because of these specific techniques. This is due to the method they chose, and a more focused, follow-up research would need to be conducted to answer this question. However, they did find that users who are unreflective (take very little time to answer questions) are statistically more likely to avoid or abuse hints. The high level parameters which they extracted (such as the number of exercises completed, and the scores achieved on those exercises) are still useful information for teachers as well as students to get a better understanding of the learning process. In another study [LMMP+14], a rule-based approach was taken to detect emotions (boredom, confusion, frustration, and happiness) known to be related to learning gains. The cited research eventually resulted in an extension of the e-learning platform [RVLK15] that allows teachers to better understand the learning process of their students. Finally, qualitative research was done by Murphy et al. [MGK+14] that looks at how Khan Academy is employed at different educational institutions, and what the attitude of teachers is towards such an online learning platform. The findings are predominantly positive.

Despite its success, Khan himself has received criticism regarding his approach to e-learning, most notably from the Washington Post [Str12] and former Apple designer Bret Victor [Vic12]. They state that Khan does not have the "necessary pedagogical content knowledge to design and deliver quality instruction", and that (regarding the programming tutorials) "live coding, as a standalone feature, misses the point". This criticism is useful in relation to the current research since it provides additional motivation for making decision based on evidence-based principles, as opposed to merely focusing on building a fun and engaging learning experience.

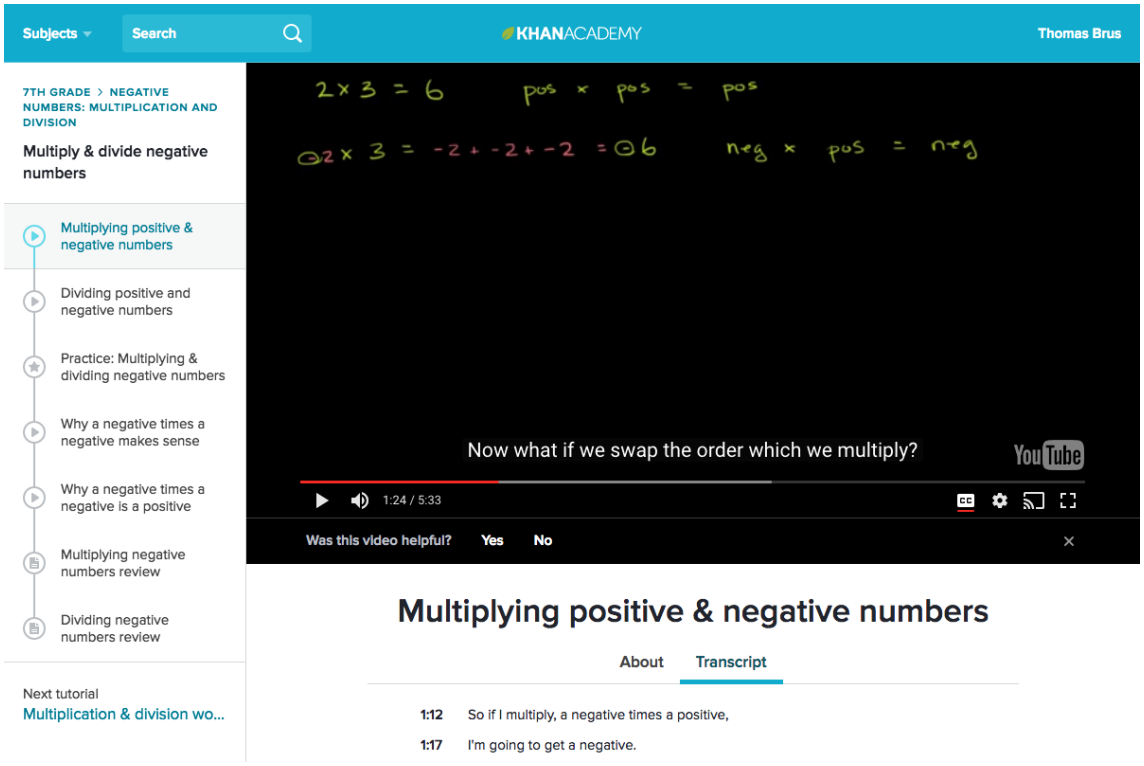


Figure 2.3.: Khan Academy. Tutorial on multiplication and division.

2.2.5. Boekhouden voor beginners

One final interesting example to review is a bookkeeping course provided by a company that also builds online bookkeeping software[4]. The course is made available for free and does not require registration. The basics of bookkeeping are addressed in logical order and are explained in articles containing mostly text. In some articles a link is made to the bookkeeping software that is offered. In other articles only specific functionality of the software is explained and the articles serves more or less as a manual page. A few topics are accompanied with an explanatory video (see Figure 2.4), however most content is designed in a static way similar to what you would find in a book. The videos show an animation that explains the given topic, supported by a voice over.

While this case study is interesting, the website is not necessarily sufficient for users of Moneybird in order to become better at bookkeeping. This is due to a couple of factors. First of all, it does not tell users how to apply the theory presented in practice, in this case in Moneybird. Secondly, there is no way for users to verify whether they understand the topics presented other than reiterating the concepts in their heads. No quizzes or exercises are offered. Interactive exercises can support the idea of learning by doing, as discussed by Clark and Mayer [CM16]. Furthermore, the information is presented in the traditional (1.0) sense of e-learning: the same material but presented on a different medium [Ebn07]. This leaves a huge opportunity unused to engage users in the material and it hinders the chance of deeper learning to occur, as a consequence of greater engagement and motivation. As seen in the previously discussed successful examples of online learning platforms, several ways to accomplish this are techniques, such as: rewarding the users with badges and unlockables, and providing a way for users to keep track of their progress and to move forward to more difficult courses. Finally, this site provides no way for users to actively

[4] http://www.boekhouden-voor-beginners.nl/inleiding/

Figure 2.4.: Boekhouden voor beginners. This article uses written text in combination with a narrated video to explain the concept of value-added tax.

participate in the process of learning more about bookkeeping. Nevertheless, the customers of this particular bookkeeping software may still get value from it, and it serves as a clever marketing tool that may attract more new customers to the software.

2.2.6. Conclusion

In this section about related work we saw numerous successful applications of e-learning and gamification. We learned that usability and visual appearance play an important role in relation to the learners' attitude. Regarding gamification, one study points out that it should never interfere with the main content (specifically "standardized tests") of the course. This does not mean learning by doing in itself cannot be effective. Duolingo takes a very research-based approach to online learning and in combination with its attractive visual design, this platform appeals to a lot of (younger) people. The case study on Babbel shows us that it can pay off to invest a lot of time in the quality and diversity of the learning material available. In a research study related to Babbel, the use of web technology to upgrade the learning experience was explored. It was again concluded that usability and appearance play an important role in the attitude of learners towards the platform. As also noted in a study discussed in section 2.2.1, it seems that negative feedback is most often received in relation to annoyances in the interface, or technical problems. From one of the studies on Khan Academy, we learn that it is difficult to describe the impact of gamification on students' motivation when no control group is used. And finally, in the last case study we identified areas of improvement for a bookkeeping tutorial by a competitor. We wish to address these areas in our own bookkeeping training.

3. Design

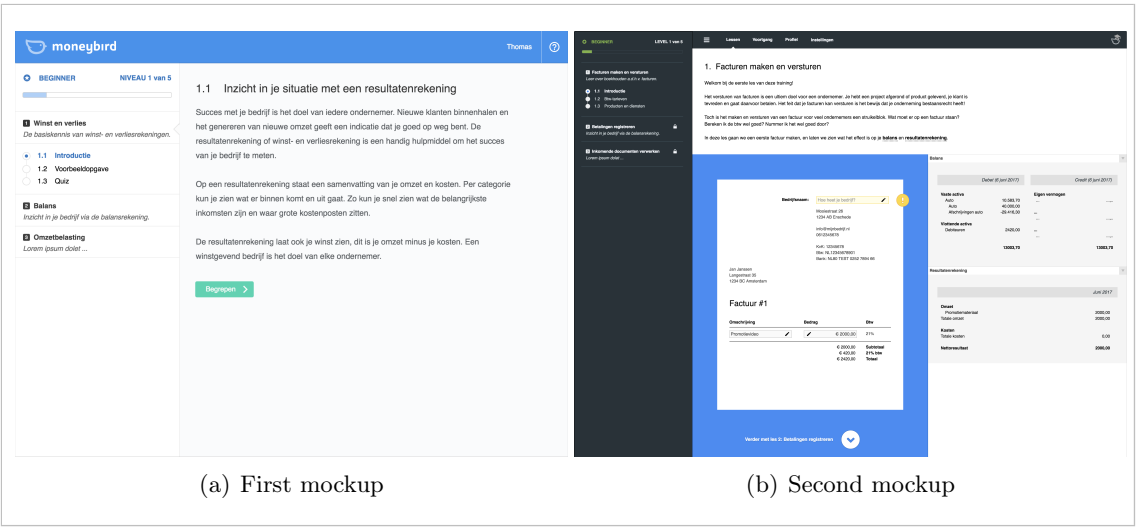
Based on the literature study that was conducted in the previous chapter, we are now ready to develop an online bookkeeping training for users of Moneybird. In this chapter we show the considerations that went into the design and how they connect to the literature that was studied.

Firstly, we chose to design the online training as a separate experience from the main application because this made it easier to only show the content that is needed for the learner to understand a given topic. When directly integrating the learning experience in the main application the user may get distracted too easily, and the focus might be too much on the functionality of Moneybird, as opposed to the bookkeeping topics we want them to learn about. Furthermore, some technical challenges presented itself in making sure that the user can only interact with the parts of the application that are needed to progress through each lesson. Additionally, we decided that it was important that learners are actively engaged in the course material via interactive exercises and multiple choice questions, similar to what Khan Academy does. Merely presenting large blocks of text is not particularly motivating for the user. This notion is supported by the main principle of e-learning: the *multimedia principle*. At the same time, it is important to strike the right balance here. Too much interactivity can distract the user and encourage behavioral activity as opposed to cognitive activity, the kind needed to actually learn something from the content presented (as discussed in section 2.1.2). In the following sections we show how we attempted to strike the right balance between a design that optimizes learning while also using persuasive design to further motivate users to make use of it.

3.1. Visual design

We initially explored two alternate directions via static designs. Both to get an idea of the visual appearance of the learning environment, as well as how the user could interact with it. We discussed both directions internally at Moneybird and weighted the pros and cons. Once we had a clear idea what the layout of the application would look like and how the different parts would come together, we started with the technical implementation. The initial idea was a very straightforward layout presented in Figure 3.1 (a). It was not yet clear what different types of content would be presented, nor whether it would just be text or perhaps something more interactive. Meanwhile, we had had several brainstorming sessions about the learning material. Not only what it would include from an educational point of view, but also how it would be presented to the user. One idea that we were particularly enthusiastic about, was to present the user with live updating bookkeeping reports that change as the user makes changes to his bookkeeping.

This idea is shown in Figure 3.1 (b). As the user makes changes to the invoice on the left, the bookkeeping reports on the right would automatically updated. This is something not currently possible in Moneybird and it provides great insights for the user. However, we realized that this layout takes up a lot of horizontal space. Furthermore, we were afraid that once lessons got more complex it would become difficult for beginners to see what is going on. Visually it is also quite challenging to show what parts have changed. However, the main problem was that it was very difficult to include a sidebar in the layout as well. From the literature review study by Chumley-Jones et al. [CJDA02], we already

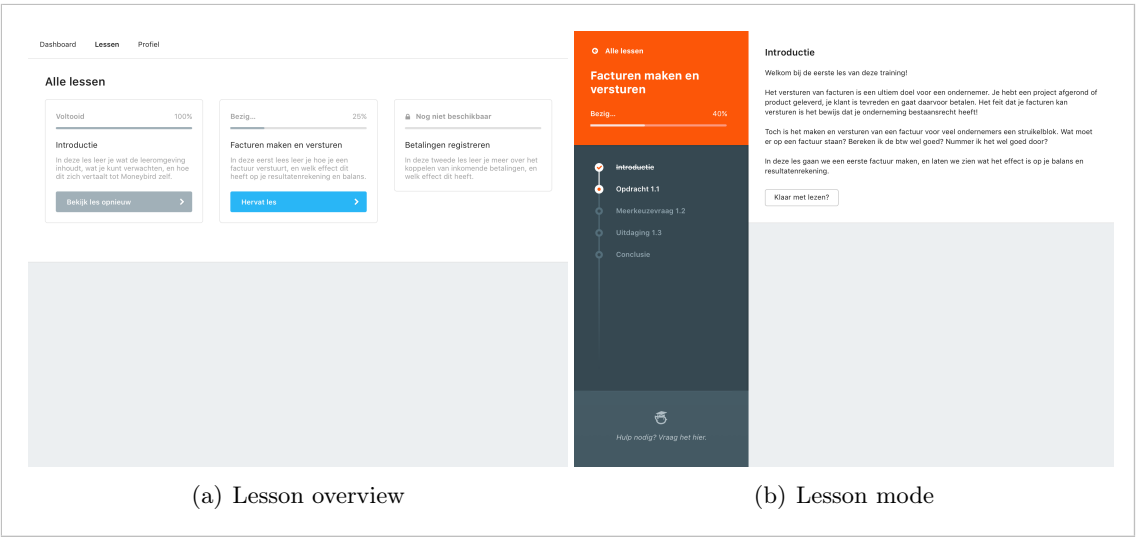


(a) First mockup (b) Second mockup

Figure 3.1.: Initial designs.

knew that having a clear navigational structure is a deciding factor in the success of an online learning environment. Since the navigation should be able to grow (depending on the number of sections a lesson consist of), and because it should ideally always be visible to the user, we decided that the sidebar should be reserved for this purpose. Due to the width of the interactive exercise and the width of the bookkeeping reports it was difficult to fit the sidebar in the layout as well.

Once it was decided that the layout in Figure 3.1 (b) would cause too many problems (practical problems such as too little horizontal space, as well as usability problems, due to the high level of interactivity), we decided to continue with direction taken in the first design. It was important that the design had a clear navigation that is always present, and that this navigation indicates the progress of the user in the lesson, as well as a means to navigate back and forth to other parts of the application. A lesson overview screen was designed as well, since the user is either choosing which lesson to follow, or is currently participating in a lesson. A clear visual distinction was made between these two modes. From the lesson screen the user can navigate back to the lesson overview screen at any point in time. These two screens are shown in Figure 3.2. The screens formed the basis for the subsequent iterations in the development of the learning environment. At this point it was already clear how the *self-monitoring principle* discussed in section 2.1.2 could be applied. This persuasive design principle states that a user with a predetermined goal should be able to rely on the system to track their progress towards that goal. This is possible via the lesson overview screen, as well as navigation and status bar displayed when in lesson mode. Self-monitoring can also include keeping track of the performance of the user. In this learning environment, the time needed to complete each section or the number of attempts taken to answer a multiple choice question have no further impact on the progress that is displayed to the user. In subsequent iterations of the design shown in Figure 3.2, the lesson material was added and various interactive exercises as well as multiple choice questions were developed. This is explained in the following sections. Later on in the research, once the usability test had been conducted, the visual design was changed once more, and the final result can be seen in section 4.5.4.



(a) Lesson overview (b) Lesson mode

Figure 3.2.: Final iteration of the visual design phase.

3.2. Lesson material

Creating the actual content of the lessons was an ongoing process. Decisions had to be made regarding the depth and scope of the lesson content as well as the order in which the content is presented. The way in which the content is presented was guided by the literature discussed in section 2.1.1. The actual content was created in collaboration with domain experts in the Moneybird team. We relied on the experience of the team in general, as well as two team members in particular: the support lead and a financial advisor (who happened to have a background in education). Both team members have a thorough understanding of bookkeeping and provided great insight in how the lessons should be structured from an educational point of view. Dealing with customer requests every day leads to a good understanding of how to best explain a given bookkeeping topic to the end user. For each lesson we identified the main learning objectives. For example, the learning objectives of lesson 1 are for the user to learn about the steps involved to send a valid invoice, and to understand what effect this action has on the profit and loss statement as well as the balance sheet. The role of taxes is purposely excluded and is addressed in lesson 2. Four lessons were written out in detail and implemented in the learning environment, as shown in Figure 3.5. Each lesson consists of an introduction, an interactive exercise with instructions and a cause-effect explanation, a number of multiple choice questions with specific feedback for each possible answer, and a summary with the most important learning points. While writing the content we tried to be concise above all, and follow a logical structure in which new concepts are introduced once they become relevant. During the usability test we collected feedback on the first three lessons. The fourth lesson was developed after the usability test had been conducted.

Type	Description	Figure
Introduction	A short textual overview of what the user is about to learn.	N/A
Exercises + explanation	An interactive simulation in which the user performs some kind of bookkeeping task, such as sending an invoice or linking an incoming transaction.	3.3
Multiple choice questions	Questions about the aforementioned exercises and the explanation that was given along with it.	3.4
Summary	Final section in which the key points of the lesson are repeated for the user to review.	3.6

Table 3.1.: The different types of content each lesson consists of.

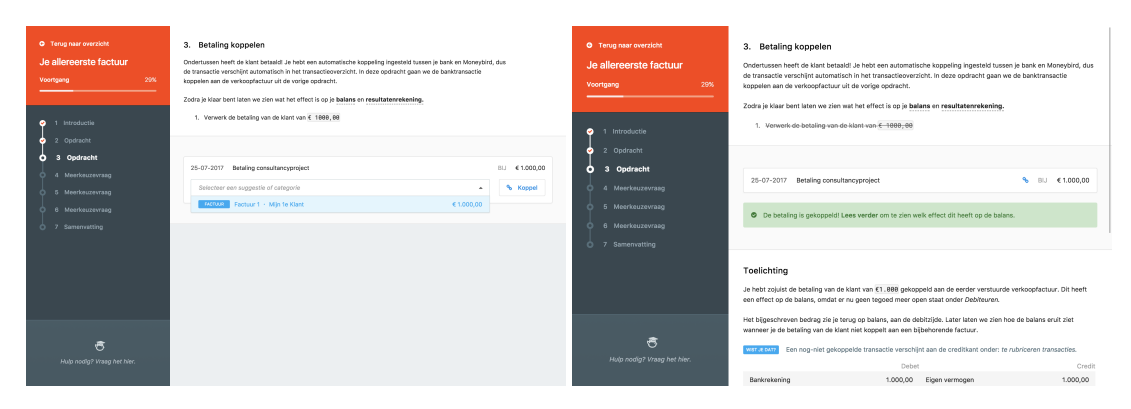
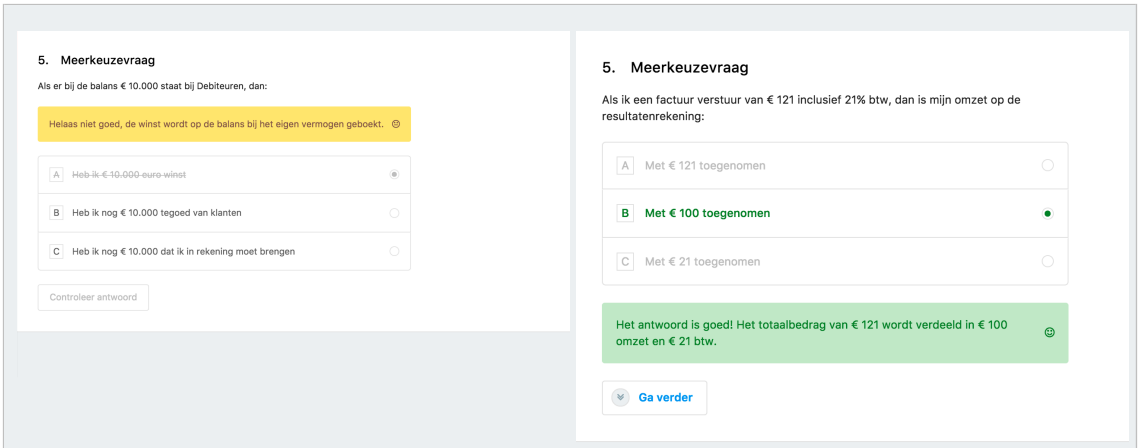


Figure 3.3.: Flow of one of the exercises in the learning environment. On the left, an exercise in which the user links an incoming transaction to a sales invoice is shown. To keep things simple, in this first exercise only a single suggestion is provided. Once the transaction is linked, an explanation follows on the effect of the action on ones bookkeeping.



(a) Multiple choice question (incorrect answer). (b) Another multiple choice question (correct answer).

Figure 3.4.: Multiple choice questions.

1: Sales invoices Introductory lesson about *sales invoices*: invoices send to a customer after selling an item or providing a service.

2: Value added taxes Follow-up lesson about sales invoices, this time with an explanation of how taxes are handled in a bookkeeping.

3: Purchase invoices Lesson about *purchase invoices*: invoices received after buying an item or service from another company.

4: Receipts Final lesson about *receipts*. Physical receipts can be scanned and uploaded in Moneybird and are processed similarly to purchase invoices, except that taxes are included in the price instead of excluded.

Figure 3.5.: The four lessons that the learning environment consists of.

3.3. E-learning principles

In this section we review the application of the various evidence-based e-learning principles in the context of the learning environment. Examples as well as suggestions for future work are discussed.

Segmentation As we learned earlier, the *segmentation principle* states that deeper learning occurs when content is presented in logical chunks with breaks, as opposed to continuously without breaks [MC01] [May05b] [CM16]. We naturally applied this principle by breaking up the content into different lessons. Each lesson consists of several sections. After each lesson is completed the user returns the lesson overview screen and can choose to continue with the next lesson once ready. Within lessons, segmentation is applied as well by having a clear separation between the different sections of a lesson. The user can navigate back and forth between completed sections if needed. At the end of each section, the user explicitly presses a continue button to complete the current section and continue

to the next one. The different types of sections are shown in Table 3.1.

Modality Note that we only made use of the visual cognitive channel to present the learning material. The addition of narrated videos as a lesson content type would have better utilized the *modality principle*, which states that deeper learning occurs when information is split across two separate cognitive channels (auditory and visual) instead. Furthermore, online platforms such as Khan Academy and YouTube further support the notion that videos are currently an extremely popular and arguably effective means to present learning material. We also felt that the videos could have helped to make the content more appealing and engaging for novice learners. It seems that including narrated videos in the learning environment was an obvious choice to make.

However, ultimately we found too many drawbacks, the main one being the production time needed to create appropriate animations or screen recordings to accompany the explanatory audio, as well as the time needed to write out the narrative and to record it in a professional way. If we had decided to go through with this idea, we would have to go all in on videos, and skip out on the other aspects of the learning environment (textual explanations, multiple choice questions, and interactive exercises) that we also deemed important. Finally, it would have been hard to make changes to the videos in the future. Nevertheless, this is still something that can be done in the future and there is nothing in the current design that stops that from becoming a reality.

Contiguity Each section is displayed on a separate page. Only for the interactive exercises, the corresponding explanation is shown on the same page. This is where the (spatial) *contiguity principle* comes into play, which argues that related pieces of information (such as graphics along with explanatory text) should be kept together as close as possible [HHH09] [May89] [MA92] [MSBM95]. Another clear example of this principle is shown in Figure 3.7 where the definition of a key term is shown right along with it instead of on a different page or in some popup window. Finally, the contiguity principle was applied in other situations where the user receives feedback on an action that was performed. For example, when answering a multiple choice question (see Figure 3.4) or when submitting a solution for an exercise. When the user submits an exercise and has entered invalid data, the error messages are shown as close as possible to those invalid values entered by the user.

Figure 3.7.: An example of how the contiguity principle is applied. Key concepts are introduced on the fly. An exact definition is available to users by hovering over the concept (as opposed to in a popup window or on a different page).

Figure 3.6.: Lesson summary. Instead of providing all information at once, the segmentation principle is applied to present each key point individually.

Coherence The relatively empty visual design with few distractions was motivated by the expected complexity of the interactive exercises that had yet to be implemented, as well as the *coherence principle* discussed in section 2.1.1 about e-learning which tells us that adding additional elements or media that are not relevant to the content presented actually interferes with learning [PMSB11] [MSA+14] [CM16]. The decision to separate the learning environment from the main application was also motivated by this e-learning principle, because existing elements in the main application would otherwise needlessly distract the user from the learning objective.

Pretraining In section 2.1.1, the *pretraining principle* was discussed as well. We did not make use of this principle, even though the research suggests that providing a quick overview of the key concepts beforehand can aid the learning process [MMW02] [KKvM06] [PCS02] [Ayr06] [CAS05]. During the usability (Chapter 4) test we learned that there were

indeed several concepts that some beginners struggled with. These were mostly technical terms used in bookkeeping, such as the names (and meaning) of different reports, as well as the categories used on those reports. We could have chosen to explain key concepts before the start of the first lesson. The reason we did not do this is that we felt the (recently signed up) users might skip over the information in the beginning since they may be impatient or they do not yet realize they will need the information later on. In a typical educational setting (when following an (online) course as part of one's education), this might be different. Thus, we chose to present the information implicitly via examples, in a way that it becomes clear from the context what each concepts entails. We felt it was important to capture the attention of the user from the very start. Instead of having them first read about these (perhaps still meaningless) concepts, we chose to let users get started with a very simple exercise first, and then gradually explain the effects on their bookkeeping as a result of completing that exercise (for example: sending their first invoice). However, to still provide some guidance to users in finding out what a certain terms exactly means, inline definitions of key concepts were added that can be revealed by hovering over them, as shown in Figure 3.7.

Personalization Finally, when discussing the writing style to use, the *personalization principle* (as discussed in section 2.1.1) was taken into account. This principle states that deeper learning occurs when the learner experiences a stronger social presence, through conversational style or via the presence of a virtual coach [CM16] [Wik17a]. Cognitive research on discourse processing shows that people put more effort in understanding material when they (feel like) they are in a conversation, as opposed to when simply presented with information [BMS+96]. For beginners, an informal conversational style instead of a formal style is preferred [MM00] [Kar10]. We adapted our writing style to use informal pronouns ("jij/je" in Dutch) instead of the formal pronoun ("u" in Dutch). Furthermore, each lesson follows a pattern in which a sort of story is told that unfolds over time. In this story, the user is provided with instructions on what to do next (i.e. complete a certain exercise), and an explanation is given regarding the effect of action taken on the (virtual) bookkeeping of the user. Subsequent exercises then build on the outcome of the previous ones. This feedback cycle in which the user performs an action (such as completing an exercise or answering a multiple choice question), and then receives feedback on their performance, should increase the feeling of conversation between the user and the system. The intent here is that the application is more than a simple interface, and instead initiates a conversation with the user that advances as the user performs certain actions. Communicating feedback through a *virtual coach* to further enforce a social presence is something that was considered, but was unfortunately not feasible due to time constraints.

3.4. Persuasive design

We will now review the application of persuasive system design, as discussed in section 2.1.2. For each category of persuasive system principles, we discuss the principles that were used and how they were used to further encourage users to use the learning environment. If a system principle was not used, we discuss why, and how it could have been implemented.

3.4.1. Primary task support

This persuasive system design category refers the main activities the user carries out in the persuasive system. We discuss the application of each of the three selected design principles, as seen in Table 2.2.

Tailoring Tailoring is a key persuasive system principle that we took into account when designing the learning environment. Since we had a clearly defined target group (new users of Moneybird, presumably beginners in bookkeeping), we made sure to suit the level of the lessons as well as the length and the tone of voice to that specific audience. We took special care to be on the safe side when it comes to complexity since learning about bookkeeping can be overwhelming for beginners.

Self-monitoring The self-monitoring principle is one of the motivations for showing progress in the learning environment. In the lesson overview screen, the progress for each lesson is shown, and within a lesson the user can see exactly which sections she has completed and which sections are left.

Simulation This persuasive system design principle states that providing simulations can persuade users by enabling them to immediately observe the link between cause and effect. This immediately caught our attention of course since understanding the link between cause and effect is such an important part of bookkeeping. We effectively used simulation in each interactive exercise. Ideally, there should be multiple solutions possible to solve an exercise and the explanation that follows should differ depending on the data which the user entered.

3.4.2. Dialogue support

Dialogue support revolves around the communication between the system and the end user. The feedback should persuade users to achieve a predetermined goal (such as completing a lesson). Five aspects are discussed, as seen in Table 2.3.

Praise Praise is used on a few occasions to further encourage users to proceed completing lessons in the learning environment. Specifically, praise is given when an exercise is completed, a multiple choice question is answered correctly, and when a lesson is completed. The praise is not personalized to the specific user but general instead and along the lines of *Well done! This is the correct answer.* (see Figure 3.4, for example).

Rewards We considered integrating a reward system in the learning environment but ended up not doing this because of time constraints. For this to work, there would also need to be some notion of points or performance in the application. Two types of rewards that we considered early on were the use of virtual badges that can be shared on social media, and a discount on the subscription plan in the main application.

Liking According to this design principle, a system that is visually attractive is likely to be more persuasive. Priority was given to accomplishing this persuasive system principle, although whether we succeeded in this is of course quite subjective and the answer may differ from user to user.

Social role The inclusion of a virtual coach was considered as the instructions in the learning environment already follow a conversational style and the use of a virtual coach could further persuade the user to continue completing the lesson material. We felt like this would have made a minor difference because there are already so many factors at play when it comes to motivation for users to complete lessons, while on the other hand it takes quite a bit of time and effort to design such a virtual coach.

Expertise Of course a lot of knowledge about bookkeeping is present at Moneybird and this could have been further exploited to increase the confidence of the users in the quality and relevance of the learning material. A simple way to accomplish this would have been to include citations (tips) in the learning environment from people at Moneybird.

3.4.3. Social support

Social support is the final category of persuasive system design that we discuss. It concerns the incorporation of social elements to persuade a user to perform a certain action. Two aspects are discussed, as seen in Table 2.4.

Social comparison We did not make use of this principle, even though social comparison can serve as quite a powerful motivation. A simple idea would have been to implement a leaderboard where users with the best scores are published. The limiting factor (aside from time constraints) was that we did not yet have a notion of performance in the system, only completeness.

Competition This principle states that a system can motivate users to adopt a target behaviour by appealing to human's natural drive to compete. We did not make use of this principle because we felt it was important for the users to take the time to complete each lesson, instead of rushing to achieve some target objective. Beginners should not feel discouraged if they make a mistake, but should rather be given a second chance so the likelihood they push through is ultimately greater. Possible ways to incorporate elements of competition is by setting a time limit goal to a complete a lesson, or to let users earn points as they answers multiple choice questions.

This concludes our review of the application of the various persuasive system design principles. We saw that most are applied in some way or another, in our system. Still, many more different ways to incorporate these principles come to mind. It is difficult to conduct research that quantifies the impact of each of these principles since so many variables are at play. Persuasive system design was used as a guideline to create a learning environment that further encourages already motivated users to learn about bookkeeping. Ultimately, a combination of data analysis and user research is what leads to a better understanding of the effectiveness of these principles in the setting of our learning environment.

3.5. Conclusion

Let us conclude this chapter by summing up the key points of this chapter. We saw how an online learning environment was developed that will be subjected to a usability test in the next chapter. We also saw how evidence-based principles of e-learning and persuasive design were incorporated to further support our goal of increasing the bookkeeping proficiency of users of Moneybird. The development of logically structured and understandable learning material was accomplished by relying on domain experts at Moneybird. Finally, we also saw how some opportunities were unfortunately missed out on due to time constraints. The main ones were the inclusion of a type of reward system, a virtual coach, and the use of narrated videos to explain the lesson material. In the next chapter we find out whether the decisions made so far, mostly based theoretical frameworks and a good chunk of intuition, do in fact translate into a pleasant learning experience.

4. Experiment I: Usability test

In this chapter we discuss the methodology used to prepare and conduct the usability test. A overview of the most important usability problems is given, and the changes made to the design as a result of these findings are shown. The chapter ends with a discussion on how to interpret the results and how to move forward.

The purpose of this usability study is to identify problems and areas for improvement in the user interface and the lesson content, in preparation for the subsequent A/B test (see Chapter 5). During the A/B test it is no longer desirable to make any changes, as that would mean different people within the test or control group see different versions of the application. Furthermore, it is then also harder to figure out from what problems the users exactly encounter since it is not possible to observe how they interact with the interface. For this reason it was decided to conduct a usability test beforehand. Aside from this, a clear motivation for conducting a usability study was already provided back in section 2.1.1, where the importance of accessibility, a clear navigational structure, and the attractiveness of visual design, was stressed, with regard to the satisfaction of the learner. Furthermore, page load time was considered to be a very important factor; something we took into account during the technical development of the learning environment.

Before the usability test began, a reasonable amount of lesson content had been implemented and a fair bit of thought had already been put into the visual design and the user experience. At the start of the usability test, the first three lessons were finished, as seen in Figure 3.5. The lesson content is considered an important part of the usability of the learning environment as well since it is key in accomplishing the main goal of the application: increasing the bookkeeping proficiency of the user. Examples of problems that may arise with regard to the lesson content are that the content is too difficult for beginners, too much information is presented at once, the order in which the information is presented is illogical, crucial background information is missing, or the writing style is unclear. We were specifically interested in problems that hinder the user from completing the lessons (typical UX problems, such as getting stuck and not knowing how to continue), and issues in understanding the lesson content. To check whether participants understood the information presented, the test facilitator sometimes interfered and asked the participant to explain a concept in their own words. Finally, some feedback was collected regarding the attitude of the users towards the application. This is discussed in section 4.5.3.

4.1. Recruitment and participants

A total of six people were selected to partake in the usability test. The people that we intended to recruit for the usability test were people with little existing bookkeeping knowledge, from various backgrounds and age ranges, and with no or little experience in using Moneybird. Via the social media channels of the company people were asked to participate in this test and to fill out a short survey that was made available online (see Appendix A). Out of the 21 people that responded, six people were selected. In deciding the number of people to select, we took into account the complexity and scope of the design (relatively simple and relatively narrow scope), the standards to which the design needs to adhere (getting stuck would be annoying, but not mission critical) and the time available within the research to recruit participants, conduct the test, and process the

ID	Gender	Age	Bookkeeping level	Moneybird user?
T1	Female	25-34	Beginner	No
T2	Female	35-44	Beginner	No
T3	Male	35-44	Average	Yes (3-6 months)
T4	Male	18-25	Beginner	No
T5	Male	35-44	Average	No
T6	Male	45-55	Beginner	No

Table 4.1.: Participant details. *Bookkeeping level* is the self-assessed level of bookkeeping knowledge (beginner, average, advanced, expert) from the questionnaire in Appendix A.

results. Preferably the tests could be conducted on a single day and this was still possible with six sessions of one hour each. In the discussion (section 4.6) we look back at whether this number of participants was sufficient. With this selection we expected to at the very least spot any serious flaws, if present, and we expected these to be pointed out separately by different participants.

In deciding which people to select, we first filtered on the number of months of experience with the main application. It was not explicitly stated on social media what the requirements were to participate, which allowed for some flexibility on the researcher's end to decide once the responses were in. It was decided to select people with little experience with the application because this mirrors the target group of new sign ups for the subsequent A/B test, described in Chapter 5. From the survey results, coincidentally, it showed that overall these people also indicated they had little preexisting bookkeeping knowledge. People could indicate at which time during the day they were available and of course this constraint was also taken into account. With these criteria in place, a diverse selection of participants was made which is summarized in Table 4.1. The people that were selected had different backgrounds and professions, varied in age and gender, and, as turned out after conducting the tests, had various levels of computer skills. It is worth pointing out that participant T3 had a background in finances and in hindsight "advanced" would have better described his bookkeeping level.

4.2. Procedure

In preparation for the test, the selected participants were sent an informed consent form (see Appendix B) via email so they could read it and reply if they had any questions. Before the start of the usability test the form was signed by the participant and the principal researcher, and a copy of the form was handed out to participant to take home.

The sessions took place at the company's office. This location was decided to be the more reliable and efficient choice in comparison to other locations, such as the participant's home. By reliable we mean that fewer things could go wrong. A room was reserved with everything in place for the participants and the test facilitator to conduct the usability test without being disturbed. Once the participants were welcomed and were ready to begin, they were guided to the test room. Participants were seated at a desk with a standard desktop monitor connected to the researcher's laptop. Mouse and keyboard were provided

Time span	Activity	Location
00:00 - 00:10	Reception, small talk, coffee/tea	Canteen
00:10 - 00:15	Informed consent form + test instructions	Test room
00:15 - 00:50	Usability test: participant carries out tasks	Test room
00:50 - 01:00	Short discussion, wrap up, and closing words	Canteen

Table 4.2.: Session schedule.

as well. The laptop was positioned next to the desktop monitor. The webcam of the laptop was used to record the participant and the test facilitator, who was seated next to the participant. The main screen (the desktop monitor) was recorded simultaneously. Both recordings include audio as well. The recording software used (Quicktime) was configured to show mouse clicks, which made it easier afterwards to track and report the participant's actions. A local version of the application was running on the researcher's computer, which made it possible to debug and fix any technical problems in case they would occur.

For each participant, a time slot of one hour was allocated. The schedule that was set up in advance is shown in Table 4.2. In practice the actual usability tests took about 35 minutes, depending on the time that was left and the progress made by the participant.

4.3. Roles

The principal researcher took on the role of test facilitator. The primary reason for this decision was that it saved time. The time otherwise spent on preparing a co-worker to be the test facilitator was now spent on tying up loose ends and finishing the third lesson. In hindsight this turned out well, because four out of six participants started and completed lesson 3, which resulted in valuable insights. One such insight is that throughout all three lessons, relatively the most mistakes were made in the first multiple choice question of lesson 3 (two out of four participants answered incorrectly). Before the start of the A/B test, one more lesson was added. This means there is a gap between what was tested during the usability test and what is shown during the A/B test. This should not be too much of an issue since part of the findings apply to all lessons, or to a recurring interface element, and furthermore lesson 4 should not be too different from lesson 3. Another reason for taking on the role as test facilitator is that now it was straightforward for the researcher to afterwards have a short discussion with the participant about their experience and attitude towards the application. This would have been more difficult to accomplish and perhaps a bit confusing for the participant if the researcher had not been present at or before the session. A set up was considered where the researcher could follow along during the session in another room but this set up was technically difficult to accomplish. Also the issue of having to explain the set up to the participants would come up. Knowing that they are being watched from another room might put them off.

Special care was taken to avoid common issues that may occur when the researcher conducts the usability test. These pitfalls include: influencing the participant positively or negatively upon an action that is taken (albeit just via the test facilitator's body language), giving away too much information in response to a question asked by the test participant,

and, priming. A typical example of priming is apparent in a question such as: "What would you expect to happen when you click this button?". By giving away that this element is a button that can be clicked on, a usability problem may be overlooked. In this case, a better question could be: "Could you show how to send this invoice?". If the users then clicks on an element he or she was not supposed to click on, then that is a clear indication of a usability problem. The test facilitator was aware of these pitfalls and tried to avoid them. Still, there are still plenty of factors that might have influenced the behavior of the test participants and ultimately the outcome of any usability test. These are discussed further in section 4.6.

4.4. Test instructions

The participants were given the following instructions before the start of the test:

- Today you are the expert and your feedback will be used to improve the product;
- Take as much time needed to complete the tasks;
- There are no wrong or right actions, you are not monitored for your performance;
- Please try and think aloud as you go through the tasks. Explain your thought process and the decisions you make;
- The test facilitator may interrupt and ask to clarify your decisions and thought process;
- Feel free to ask questions if things are unclear. The test facilitator may or may not provide you with an answer.

Regarding the last point, an example of such a situation is apparent in usability report #19 (see Appendix, section C). Here the participant asks whether it is possible to navigate back to the previous lesson section. The test facilitator waits and observes the subsequent action taken by the participant, instead of answering straightaway. In this case it would have been especially interesting if the participant had not quickly figured out how to navigate back, because this could indicate a usability problem. Of course, if a situation were to occur where the user would get stuck, the test facilitator would help the user out and tell him how to continue. This did happen once, when a technical issue was encountered (see usability report #22). Furthermore, during the test some questions were asked by the participants in relation to the lesson content and the bookkeeping concepts presented. In general no conclusive answer was given to the questions but rather a hint was given or, most frequently, the question was repeated back to the participant. This often triggered them to think more thoroughly about the problem and sometimes go back to the previous screen to re-read the explanations.

Before the first lesson was started, users were asked briefly about their impression of the lesson overview screen, whether it was obvious for them how to proceed and if they felt like anything was missing on this screen at this point. No problems were encountered here and one positive response that summarizes the overall impression of the participants is written down in usability report #1 in the Appendix. Once the participants were ready and had no further questions, they were given control of the mouse and keyboard and the lesson overview screen was shown to them. The objective that was given to them was to try and complete the first three lessons. This was essentially the only task they had to carry out. It was made clear once again that it was not essential to complete all three lessons and we would simply see how much progress we could make given the time available.

ID	Time taken (mm:ss)	Lessons completed	Correct answers
T1	36:11	1, 2.1-2.5	3/4
T2	37:00	1, 2	6/6
T3	28:55	1, 2, 3	9/9
T4	35:46	1, 2, 3	6/9
T5	26:24	1, 2, 3	8/9
T6	26:27	1, 2, 3	7/9

Table 4.3.: Time taken by each participant and number of lessons completed within this time. Two participants indicated they were going to fill out an incorrect answer on purpose (out of curiosity). These attempts were not counted towards the total of incorrect attempts.

4.5. Results

Since all sessions were recorded it was straightforward to extract a number of metrics, such as the completion time, number of lessons completed and the number of (in)correct multiple choice question answers for each participant. These are shown in Table 4.3. As to be expected, the participants that were less proficient in bookkeeping needed more time to complete each lesson. A few mistakes were made but overall the participants did fairly well on the multiple choice questions. Overall, the participants entered an incorrect answer a total of six times, where the correct answer was entered (instantly or after one or more incorrect attempts) a total of 46 times. As can be seen in the table, not all participants could attempt all nine questions, but for the ones they did attempt, a correct answer was provided eventually. When further analyzing the data collected, the one question that jumps out is the first question of the third lesson, where two out of four participants entered an incorrect answer. Both participants made the same initial mistake. However, at this point it is hard to draw any conclusions from this little data. When collecting quantitative during a usability test, the goal is usually to compare the data between subsequent versions of a the same design, where during each iteration the feedback of the previous iteration is processed. The intend of this usability study was to collect feedback we otherwise could not have collected beforehand. This feedback is discussed in the following two sections.

4.5.1. Usability reports

Once the usability tests had taken place, the recorded sessions were analyzed and usability reports were created (see Appendix C) for each noteworthy interaction, interesting comment, and insight gained via questions asked to the participant. Of course, not every interaction was documented, in order to to keep the number of reports manageable. A total of 51 reports were extracted from the six recorded sessions. In some cases a literal transcription is given of the conversation (as to not interpret what actually happened), while in other cases the problem was simple enough to provide a short summary instead. The template used is based on a example available at Usability.gov [oHS17]. In the following section the most important findings are presented.

4.5.2. Findings & recommendations

In this section a summary is given of the problems encountered and comments received during the usability test (see Appendix C). Each finding includes a reference to the corresponding usability reports, which participants were affected, optionally some screenshots to clarify the problem, a short description, possible solutions, and the final verdict, i.e. what was actually changed in the end. All participants were able to progress through the lessons without encountering any major problems, without getting stuck, and without becoming agitated or annoyed. There was one technical issue with the multiple choice questions which was a consequence of some technical changes that were made in preparation of the usability test. This was quickly resolved by the test facilitator.

A. Vibrant orange/red color in sidebar header stands out too much	
Screenshot(s)	Figure 4.1
Relevant usability reports	#2 #30
Participants affected	T1

Description The participant points out twice that the color stands out too much, and she indicates that it might have made sense for the color to change to green once the lesson is completed.

Recommendation Another color could be chosen that is less distracting, such as dark gray blue, or some shade of gray. However, the purpose of the color is to stand out and to visually indicate the user has entered "lesson mode". Another solution would be to choose a different shade of orange that works better across different screens and color settings.

Outcome In the end it was decided to change the colors of the layout altogether, in order to be more consistent with the visual identity of the main application. In the final design the lesson header is less distracting.

B. Confusion about purposely simplified invoice in lesson 1	
Relevant usability reports	#8
Participants affected	T2

Description The participant has preexisting knowledge about what an invoice should look like and is confused about the missing details of the invoice in the exercise. These include: the lack of a tax rate, (apparent lack of) invoice number, and lack of payment instructions near the bottom of the invoice.

Recommendation Leaving out the tax rate was a conscious decision made to keep things simple in the first lesson. Including the concept of taxes increases the length and difficulty of the explanations that follow. However, it does seem that removing things which users already expect to be present, in order to simplify, can be counter-effective with regard to how intuitive the interface is.

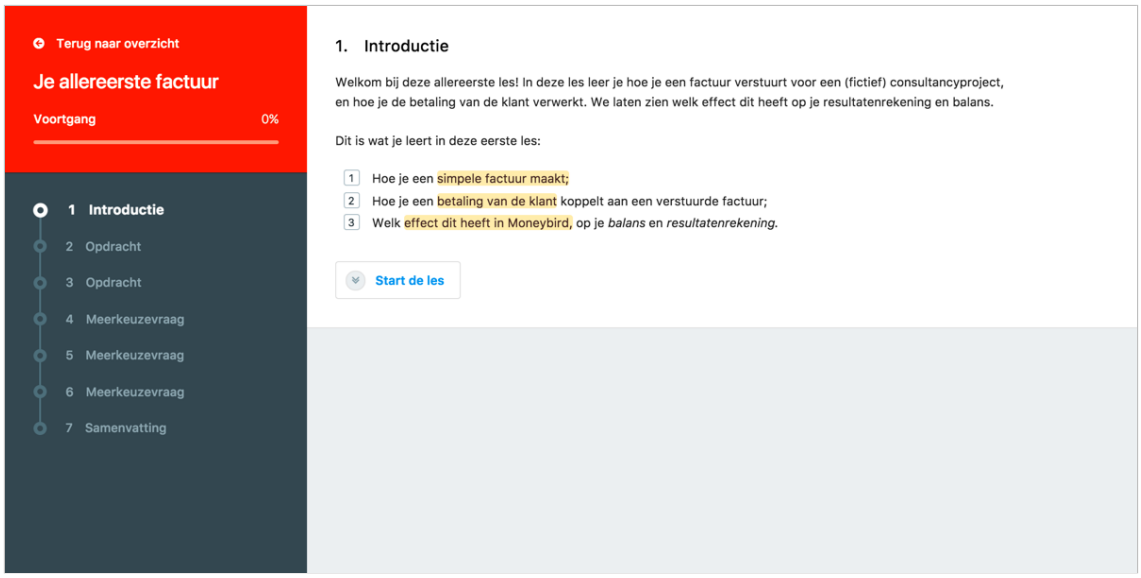


Figure 4.1.: A participant indicated the header stands out too much in comparison to the rest of the design.

Outcome It was decided to not make any changes for now, for two reasons: only one participant was affected, and it requires time and effort to implement the changes without introducing new problems.

C. Not instantly obvious that the user should scroll down after completing the invoice exercise

Screenshot(s)	Figure 4.2 and 4.3
Relevant usability reports	#8 #9 #11 #13
Participants affected	T2, T3, T4 T5

Description Several participants indicated that it was not immediately obvious that they should scroll down after completing the invoice exercise. Depending on the current position of the scrollbar, a green message was shown near the bottom of the screen. This message indicates that the user should scroll down to read the explanation and to continue the lesson. The height of the web browser was purposely set to highlight this potential issue. During the test this resulted in some slight moments of confusion and a number of comments. Regardless, all users could proceed without further issues.

Recommendation It is relatively safe to conclude that this issue will not occur when the screen height is sufficient for the success message to be in the viewport. It is hard to say how often this will be the case. Thus the preferred solution here would be to use sticky footer. This solution works as follows. When the success message is below the bottom of the viewport, it is repositioned to stick to the bottom of the screen, similar to how the send button and invoice status label (see Figure 4.3) stick to the top of the screen. When the success message *is* within the viewport, it retains its original position within the layout.

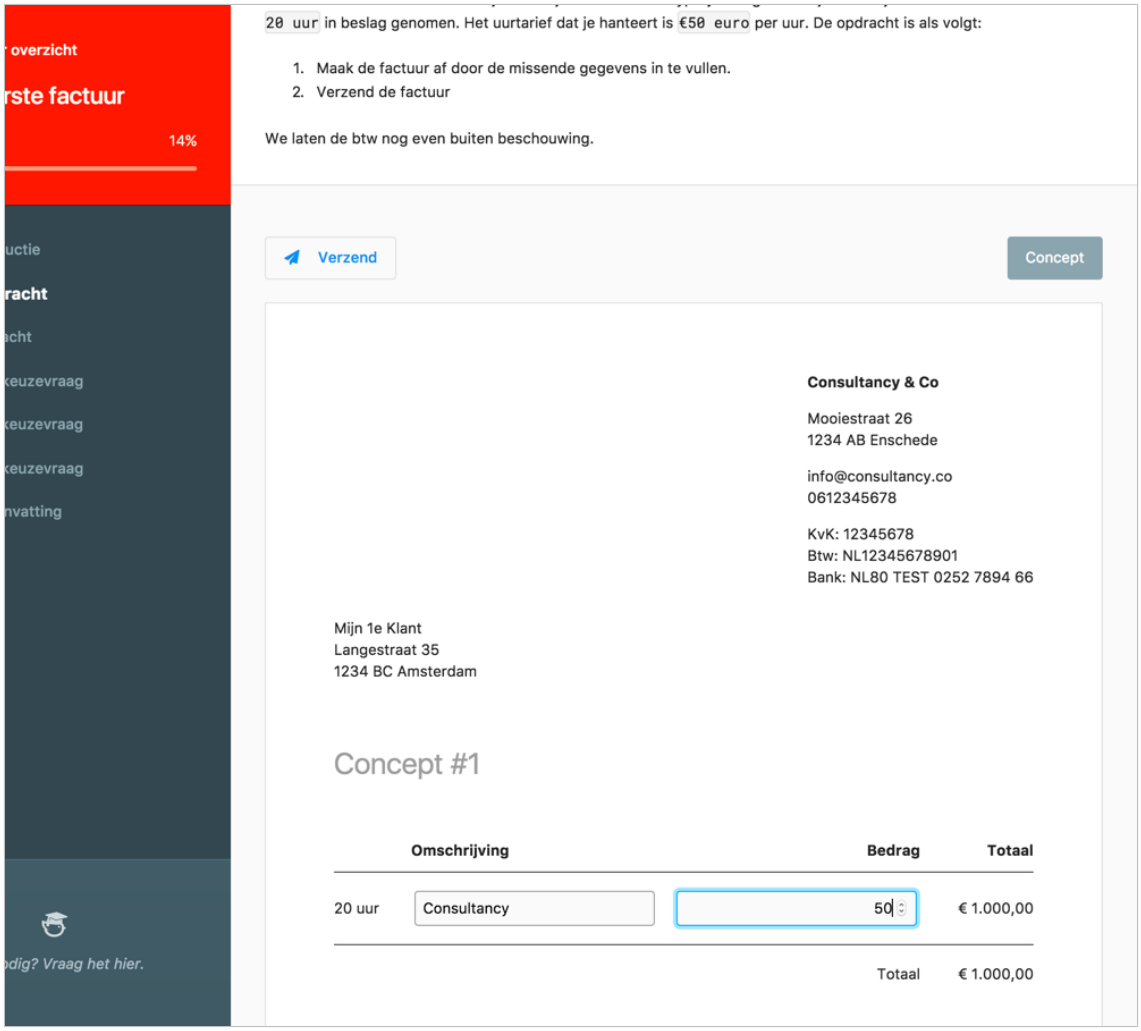
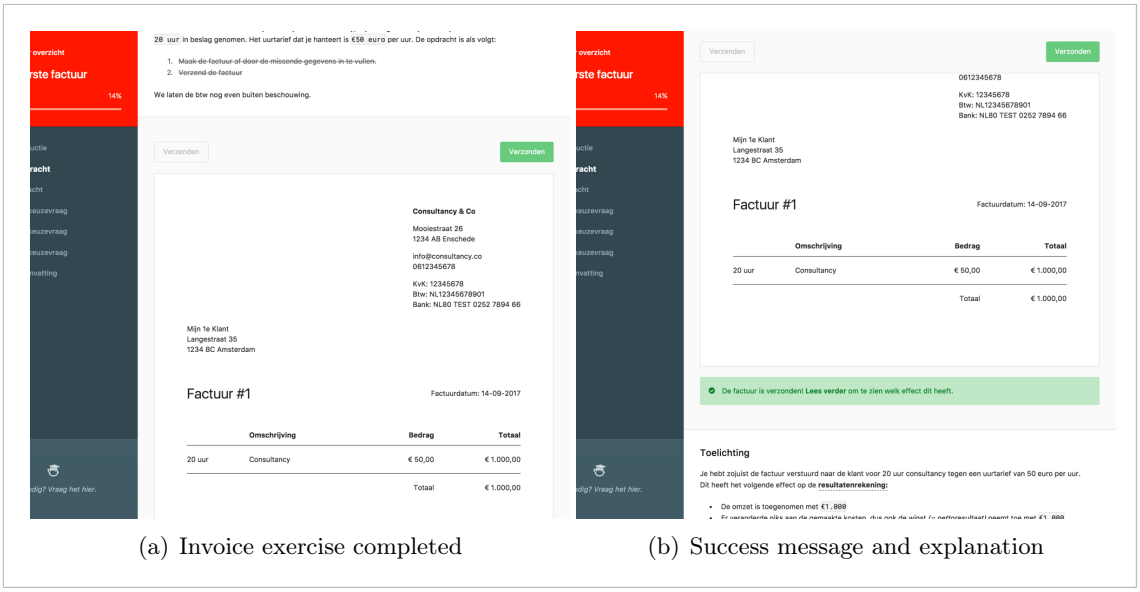


Figure 4.2.: The participant completes the invoice exercise by entering the correct data and pressing the send button. Via an animation the invoice transitions into the completed invoice shown in Figure 4.3.

Outcome The sticky footer as mentioned above was implemented.

D. Mismatch in expectations regarding position of send invoice button	
Screenshot(s)	Figure 4.2 and 4.3.
Relevant usability reports	#7 #34 #36
Participants affected	T1, T2, T4

Description Several participants pointed out that they had expected the button that sends the invoice to be positioned below the invoice (instead of at the top left corner). One participant also pointed out that it would make sense if the button was positioned below and near the right. Her explanation for this suggestion was that she started scanning at the top-left corner and when she was done filling out the details she ended up at the bottom-right corner.



(a) Invoice exercise completed (b) Success message and explanation

Figure 4.3.: Once the invoice exercise is completed a sound is played, but a success message may not be visible until the user scrolls down.

Recommendation The position of the send button has been a topic of discussion since the very first mockups and in earlier designs the button was even position centered below the invoice and at one point right aligned below the invoice. However, in order to keep things consistent with the main application, the send button is now positioned at the top left corner and the invoice status label is positioned at the top right corner. The one participant that had previous experience with Moneybird (3-6 months) made no comment about the position of the button and proceeded effortlessly (#35). Since the button and the label use sticky positioning, they stick to the top of the screen when the user scrolls down, which guarantees that they are always visible when the user is done filling out the input fields.

Outcome No changes were made to the position of the send button, despite the number of people affected. For now, consistency with the main application is considered to be of more importance.

E. Difficulty in deciding how to continue after answering a multiple choice question incorrectly

Screenshot(s)	Figure 4.4.
Relevant usability reports	#26
Participants affected	T1

Description This was the only moment during all of the test sessions where a participant really had difficulties on figuring out how to continue. As shown in the screenshot, after answering a multiple choice question incorrectly, it is not immediately obvious that one of the two other options can be selected. The submit button is also disabled.

Recommendation Part of the fix was pointed out by the participant (and, in hindsight, a mistake on our side): the text "please try again" was missing at the end of the

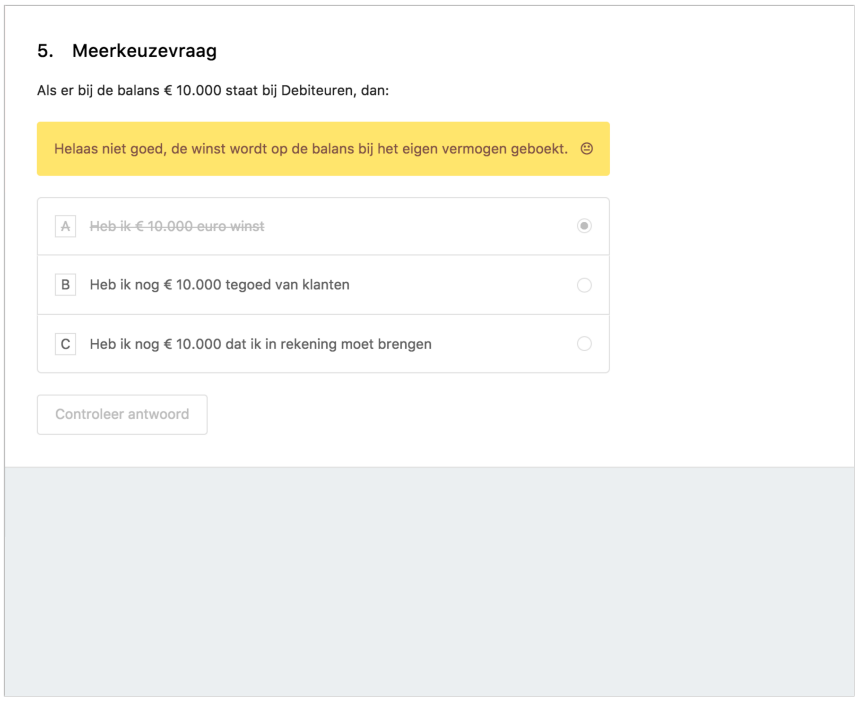


Figure 4.4.: Multiple choice question after submitting an incorrect answer.

incorrect attempt feedback text. The other part of the fix would be to create greater visual contrast between the enabled and disabled state of answers.

Outcome The missing text was added, but no further visual changes were made since only one participant was affected.

F. Confusion on key concept explained in summary of lesson 1

Relevant usability reports	#29
Participants affected	T1

Description One participant points out she does not understand one of the key points of the summary, namely why a profit is made the moment an invoice is sent. She is confused about why this is the case since the client may end up not paying, leaving her with no profit at all. So she indicates it would have made more sense if the term "expected profit" was used. This misunderstanding is common among novice users and can be explained as follows. The issue here is that *revenue* is booked whenever an activity is carried out, regardless of whether the client has paid yet. Since *costs* stay the same, a profit ($revenue - costs$) is made the moment an invoice is sent.

Recommendation Only one participant pointed out this specific problem, yet it is quite an alarming issue and it is likely other novice users might be wondering about it as well. The recommendation here is to review whether the concept is explained clear enough and then to decide whether it is worth additional explanation, or whether it needs to repeated more than once throughout the lesson.

Outcome Due to time constraints no further changes were made to the lesson content regarding this specific issue.

G. Possibly annoying remark in summary of lesson 1

Relevant usability reports	#32
Participants affected	T5

Description The participant points out that the sentence "that wasn't so difficult, yet, right?" might annoy users that did have trouble completing the lesson. He assumes that the text is not dynamically generated depending on the performance (which is correct) and he points out it might be wise to tread lightly here.

Recommendation The quick solution is to rewrite the sentence to something less presumptuous. For the long term, it may be worthwhile to dive into the research of adaptive learning and figure out how feedback influences the motivation of the user and whether it should be shown dynamically so that it fits the actual performance of the user.

Outcome The sentence was changed to something more friendly ("hopefully this was easy enough to understand so far").

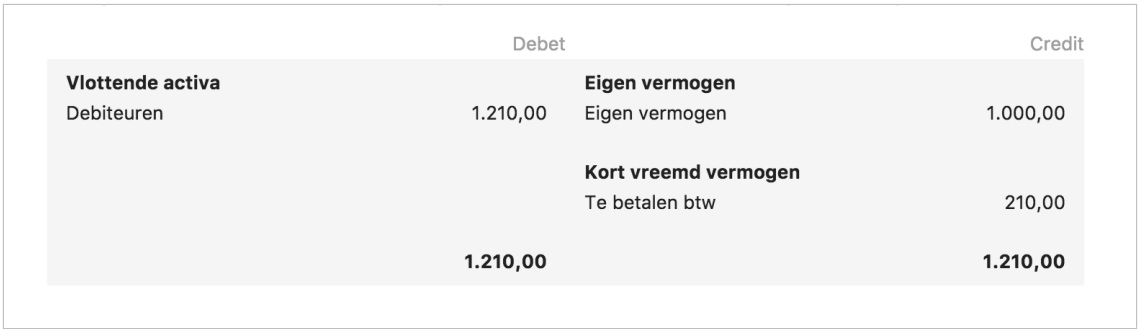
H. Suggestions for improvement regarding showing differences between reports

Screenshots(s)	Figure 4.5 and 4.6.
Relevant usability reports	#38 #39 #49
Participants affected	T2, T4

Description Two participants experienced some trouble in indicating the difference between two balance sheets presented on subsequent pages within a lesson. In this scenario, only a single line would change ("Debitors" became "Bank account"). It was sometimes hard to notice this difference, even though participants could deduce from the exercise that the reports had to be different. Several suggestions were made to improve this situation. One suggestion was to highlight the difference in a special color. Another was to show both lines even though one of them is redundant. And finally, one participant suggested to show the previous balance sheet before the current balance sheet, but collapse it by default. The user can then expand it to compare the two reports.

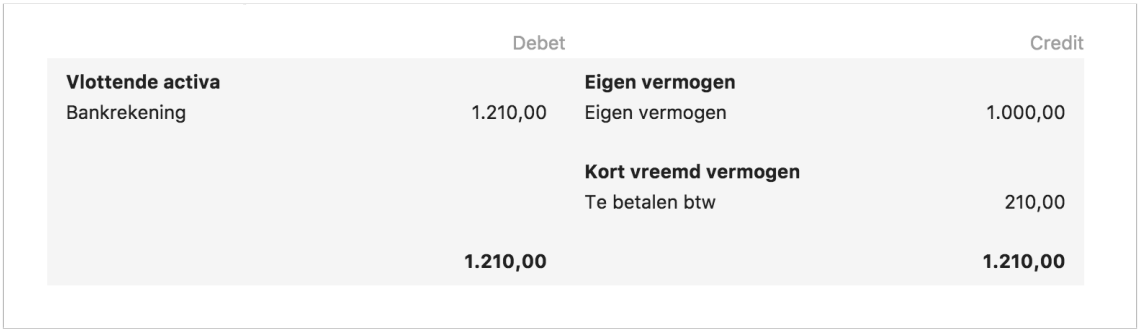
Recommendation As the reports become more complex, and the differences become smaller in comparison, a solution needs to be found for this issue. A mockup of such a solution is shown in Figure 4.7.

Outcome No changes were made because of time constraints, however, the idea presented in the mockup is something we do want to introduce in the future.



	Debet		Credit
Vlottende activa		**Eigen vermogen**	
Debiteuren	1.210,00	Eigen vermogen	1.000,00
		Kort vreemd vermogen	
		Te betalen btw	210,00
	1.210,00		**1.210,00**

Figure 4.5.: Balance sheet after sending an invoice (lesson 2, section 2).



	Debet		Credit
Vlottende activa		**Eigen vermogen**	
Bankrekening	1.210,00	Eigen vermogen	1.000,00
		Kort vreemd vermogen	
		Te betalen btw	210,00
	1.210,00		**1.210,00**

Figure 4.6.: Balance sheet after the client has paid, and the transaction is linked to the invoice (lesson 2, section 3). Notice there is only a small difference in comparison to Figure 4.5. This difference can be hard to spot in between two pages when a lot of other information is present that also requires the user's attention.

	Debet		Credit
Vlottende activa		**Eigen vermogen**	
~~Debiteuren~~	~~1.210,00~~	Eigen vermogen	1.000,00
Bankrekening	1.210,00		
		Kort vreemd vermogen	
		Te betalen btw	210,00
	1.210,00		**1.210,00**

Figure 4.7.: A possible solution that visually indicates the changes between two balance sheets. This balance sheet could be shown instead of the balance sheet in in lesson 2, section 3 (Figure 4.6).

I. Comments about drag and drop exercise	
Relevant usability reports	#45 #46 #47
Participants affected	T4, T5, T6

Description Two participants made a comment about the drag and drop interaction in the first exercise of the third lesson. A third participant explained his thoughts on the interaction after the researcher asked him about it. The problem here is that it was not immediately obvious for participants that something had changed after dragging the purchase invoice into the drop zone. Participants would scroll down intuitively but a bigger visual change was expected according to the participants. One participant noted that he was unsure about the added value of dragging and dropping the attachment. The idea here is that in the main application, users can drag and drop an invoice they received from a client, after which a preview is generated and some details are already filled out in the purchase invoice form. In this exercise the idea was to simulate this behavior.

Recommendation Upfront it was clear that this interaction needed some polishing, so it is no surprise these comments were made. The most straightforward change to implement is to visually change the drop zone once the attachment is handled so that it is clear something has happened. An instructional text could be placed, telling the users to scroll down. The one comment about the interaction being redundant is interesting but more evidence is needed to conclude whether this is in fact the case.

Outcome The solution chosen was to replace the drop zone (after the attachment is dragged into it and while the attachment is being processed) with the preview that is generated, instead of showing both underneath each other. With this new solution it is always obvious that something has happened, since the drop zone was already within the viewport, thus the preview area is also (at least partly) within the viewport as it replaces the drop zone.

Finally, there were some minor additional issues regarding the lesson content. One participant had trouble answering a multiple choice question due the word choice (usability report #28). It was unclear that "debitors" referred to a category on the balance sheet as opposed to actual (multiple) clients. This issue was fixed by changing the wording. Furthermore, the participant noted he felt like a lot of content was presented at once at one point (usability report #37), and that one option he did expect to see was missing from a multiple choice question (usability report #50). These issues have been investigated, but no further changes were made because the problems were not considered to be severe enough to justify the investment in time and effort to implement the changes.

4.5.3. Positive aspects and compliments

The following spontaneous comments are reported that point out positive aspects of the design and lesson content: #1, #3, #10, #15, #16, #21, #31, #33, #40, #43, and #51 (see Appendix C). This provides further assurance that the overall design is on point. No problems were encountered on the lesson overview page and it was straightforward for participants to continue from one lesson to another. The lessons could be completed relatively straightforward and no interference was needed from the test facilitator. In almost all situations the participant carried out the correct subsequent actions intuitively and efficiently.

It was interesting to see that participants responded positively to the smileys and sound effects used (see Figure 4.8); we conclude these are welcome additions to the application. Overall positive feedback was received from participants towards using the application in the future. One participant said he would, but that he would prefer to watch short videos instead. Another participant was very enthusiastic and asked whether the lessons would be available for free or whether she would need to pay for them. Yet another participant showed great interest in the learning environment given more advanced topics

are addressed. Not all participants were asked explicitly whether they would use the application in the future.

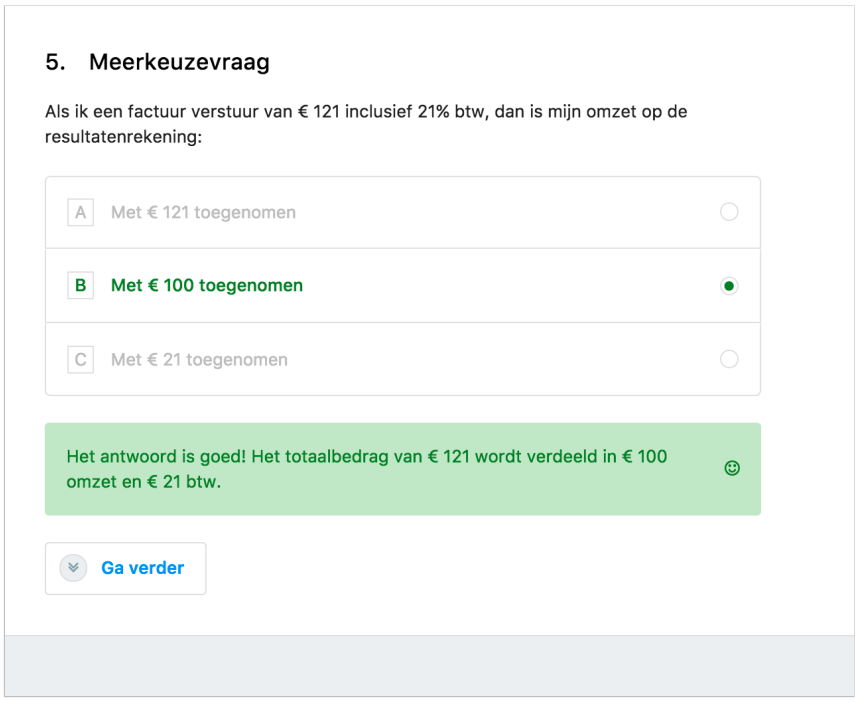
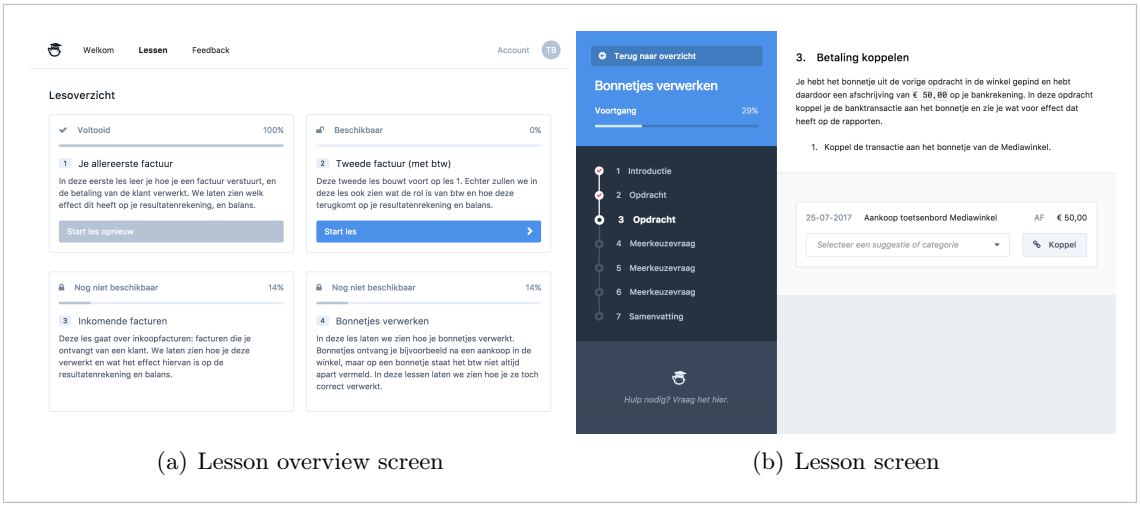


Figure 4.8.: Visual as well as auditory feedback (via sound effects) is provided throughout the course as users complete exercises, answer multiple choice questions, and finish lessons.

4.5.4. Additional changes

Aside from the changes made as a result of this usability test, we also made some final changes to the visual appearance (i.e. the color scheme) of the learning environment to ensure a more consistent user experience across the two applications (learning environment and the main application). The final design is shown in Figure 4.9. We made sure to maintain visual contrast and to use similar colors as much as possible, since colors often carry specific meaning. We evaluated internally whether the changes introduced any new usability problems. We assume the changes introduced no major new usability problems, and instead provide a greater feeling of coherence and recognition for users coming from the main application to the learning environment. The final changes were made after the usability test was over. Once the A/B test (Chapter 5) had started, no more changes were made.



(a) Lesson overview screen

(b) Lesson screen

Figure 4.9.: Final design of the learning environment.

4.6. Discussion

Before we come to the conclusion of this usability study, let us make a few remarks regarding the methodology used. As pointed out in section 4.2, the decision was made to conduct the usability test at the company's office. It can be argued that ideally the test would have been conducted at the participant's home or workplace, since that is the environment in which they usually do their bookkeeping. On the other hand, with Moneybird it is possible to do online bookkeeping from anywhere, so in that regard the location should not make too much of a difference. From a practical point of view it was not feasible to travel to each participant and to do all the extra preparations needed to make sure the test would take place without problems.

On the other hand, a much bigger concern is the influence of a test facilitator being present, the fact that these participants were self-selected, and that they were compensated financially. It is likely that the participants put more effort in understanding the lesson content and answering the multiple choice questions. It is also possible that participants responded more positively to the interface because of the presence of the test facilitator, although that seemed not to be the case. The possible influence of the test facilitator was discussed in section 4.3. However, the fact that they are self-selected does play a role in the type of feedback that can be expected. In general self-selected participants are more outspoken and most likely more enthusiastic than the average user. In this case the participants were mostly unfamiliar with the main application, which did work in our favor, since the participants were not specifically biased in a positive or negative way towards the application and the brand.

We are satisfied with the sample size since it provided a nice balance between the amount of data and feedback collected, and the time and energy needed to conduct the usability tests. However, in a fair number of cases a problem would be named only once, by one participant only. In those cases it is hard to draw a conclusion on whether action should be taken or not. It can be difficult to gauge whether the participant represents an average user and thus the problem will be encountered more frequently in the future, or whether the participant is simply an outlier and the comment comes down to personal preference.

Finally, not every participant was asked explicitly after each and every lesson whether they understood the content completely. This means it is possible that some participants had a certain understanding of what it was they should have learned but that this understanding was incorrect. In most cases it seemed obvious to the test facilitator that

the participant understood the content, judging from the comments made and the answers provided to the multiple choice questions, as well as the way the participant would interact with the application. However, since understanding the lesson content is such an important part of the overall application, this aspect could have been examined more thoroughly by the test facilitator.

4.7. Conclusion

The main goal of this usability test was to discover any major flaws if present in the interface, so that they could be fixed before the start of the A/B test. We were also interested in which parts of the application needed improvement or did not yet work as intuitively as we desired. Finally, an additional goal was to gain insight in how well the participants understood the lesson content and whether they would actually have learned something by using the application.

Regarding the usability of the application, the conclusion is that no major flaws were found and that the participants could proceed through the lessons without help or interference from the test facilitator, except for one situation in which a technical issue was encountered. This issue was quickly resolved by the test facilitator. Aside from this, no major issues were encountered. A fair number of minor issues were found though, with the most important ones summarized in section 4.5.2. Especially the issue where it was not instantly obvious how to continue after completing the invoice exercise, came up numerous times. This issue was resolved by using a sticky footer so that the success message is always within the viewport. What was especially useful about conducting the usability test, is that we now have a better idea of whether the content presented is understandable for novice users. It turned out that most test participants could understand the content fairly well. The two participants with presumably the least bookkeeping knowledge (T1 and T4) showed some insecurities and hesitation during the test and experienced some difficulties in answering the multiple choice questions. This might be an indication that the lesson content needs to be explained more clearly, or in more steps. It could also be an indication that some background concepts are not explained well enough, such as what a balance sheet is and what a profit loss report entails. However, overall, especially in combination with the positive findings summarized in section 4.5.3, we are confident that after having implemented the suggested improvements, the application is ready to be used in the A/B test.

5. Experiment II: A/B Test

In this chapter we explain how the final experiment of this research was conducted; an online experiment that includes two variations of the main product, one with a reference to the learning environment and one without. This experiment is a type of statistical hypothesis testing that is commonly known as an A/B test. Users are randomly split up in two groups: the control group and the test group. Since users are not aware they see either one of the versions, a possible difference in behavior between the two groups can be fully attributed to the modification that was implemented and shown (or no modification in case of the control group). Companies that conduct A/B tests typically decide between two versions of a design by relying on metrics such as conversion (the number of users that upgrade to a paid plan after a given period of time), or the number of times the user performed a certain action, such as sending an invoice. Depending on the probability distribution of these metrics, an appropriate statistical test is chosen to decide whether the observed change between the two groups is statistically significant.

The goal of this experiment is to study whether the learning environment, as designed during the course of this research, influences the bookkeeping behavior of users in the main application. Let us revisit the two sub-questions posed in section 1.3:

1. *What conclusions can we draw on the effect of the online training in relation to bookkeeping proficiency?*

2. *What conclusions can we draw on the effect of (the presence of) the online training in relation to user engagement?*

With this experiment we aim to investigate whether offering an online learning environment to new users of Moneybird influences the behavior of users in the main application itself. To answer the first research question, we investigate whether the actions users perform in the main application indicate improved bookkeeping skills and better understanding of bookkeeping concepts. In other words: increased proficiency in doing bookkeeping using Moneybird. To answer the second sub-question, we measure user engagement in the main application. This denotes the frequency and quality of interactions between the customer and the product. In the next section we explain the exact approach that was taken to carry out the experiment.

5.1. Methodology

In this section we describe the steps taken in designing this experiment. We explain how to exactly select new users for the A/B test, how long they should be part of the experiment, what metrics we collect, and finally how to analyze the collected data.

5.1.1. Selection criteria

Recall the decision in section 1.2 to focus on the target group of new users. In this section we describe the target group more precisely. First of all, newly signed up users are split in two groups based on their pseudo id. A pseudo id is a randomly assigned id that each user has and that is stored in the event logs. The control group consists of new users with an *odd*

pseudo id and the test group consists of users with an *even* pseudo id. The control group is used as a reference point to compare against and decide whether observed changes are statistically significant. This is explained in more detail in section 5.1.5. The test group is the group that is invited to partake in the learning environment. It is important to precisely define the A/B testing period, as well as which users are part of the test group and the control group, and which users are part of neither. This is because it is possible for *any* user to access the learning environment, given they know the URL. So we do not want to include users in our test group that did use the learning environment but do not fit the aforementioned criteria. The same goes for the control group. Furthermore, we want to limit the number of people that are included in the experiment. This was done by limiting the period in which users can be selected for the A/B test. In section 5.1.3, a more detailed explanation is given. Finally, we decided to keep showing the onboarding changes (as explained in section 5.1.2) even after the experiment is over, to ensure a consistent user experience. For the sake of the experiment we measure only the behavior of the users during the testing period, as to not run the experiment indefinitely.

When constructing exact selection criteria we first decided that we only wish to select Dutch users, since the learning environment is available in Dutch only. In Moneybird, users can set a preferred interface language, so we simply use the value of this setting to filter out non-Dutch users. Users can gain access to Moneybird in multiple ways, for example via a developer account or by being invited by another user. We only want to select users that signed up by themselves via the website. This way we can be reasonably sure users are completely new to the application. Of course they could have previously signed up with a different email, but this scenario is very hard to detect. And finally, the user needs to have signed up in a specific date range (see section 5.1.3). For the control group the criteria are the same except that we select users with an *odd* pseudo id. With these requirements in place the onboarding process for the selected users could be implemented. This is explained in the following section.

5.1.2. Onboarding flow

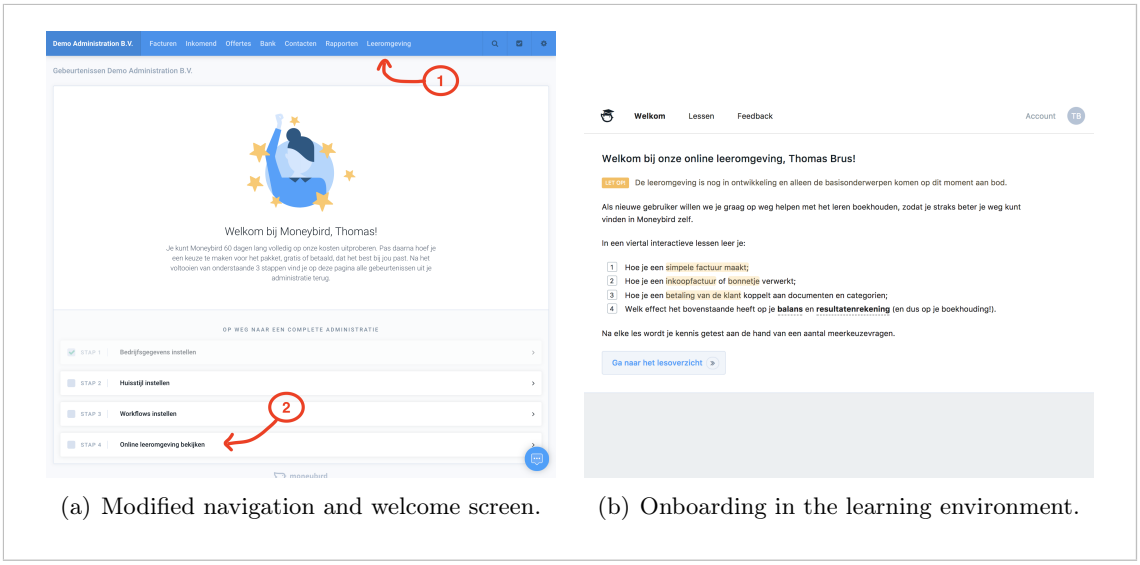
The moment a user in test group visits the main application, we are faced with two challenges:

1. They need to be made aware of the existence of the learning environment;
2. They need to be able to effortlessly navigate back to the learning environment from the main application.

The second point is important because it would make for a bad user experience if the user was given a link to the learning environment only once, and would then need to remember or bookmark it in order to easily navigate back to it later on. Since it is our goal to encourage the use of learning environment, both challenges need to be addressed. Two changes were made in the main application to make users aware of the existence of the learning environment, and to ensure that it is easy for users to navigate back to the learning application from the main application, at any point in time (see Figure 5.1):

1. A menu item with a link to the learning app was added to the main menu;
2. An additional fourth step was added to the onboarding screen. This screen automatically disappears when the first three steps have been completed.

With these changes in place we expect most users to discover the learning environment and visit it at least once. However, there are some limitations to this approach. Given the



(a) Modified navigation and welcome screen.

(b) Onboarding in the learning environment.

Figure 5.1.: Onboarding flow. An additional menu item (1) was added in the main application as well as a fourth onboarding step (2), both redirect the user to the learning environment (b). In the learning environment, users are greeted with a welcome message to provide some context, after which they can continue to the lessons overview screen.

type of experiment we conduct (an A/B test), it is mandatory that users are not aware they partake in an experiment. This limits our options when it comes to encouraging users to visit the learning environment. With the changes mentioned above, it is perfectly possible that only a small percentage of the target group actually visit the learning application. This in turns means that the expected effect is small.

It would be a mistake to reject all users of the test group that did not make use of the learning application. First of all, it is difficult to draw a clear line. For instance, users can visit the learning application once but still not partake in any lessons. Or they can start with one lesson but not finish it. It is hard to say what the exact new criteria then should then be for the narrowed down test group. But most importantly, by filtering on users that made use of the learning application, we unintentionally bring a *third variable* into play. In a way we implicitly divide users into a group that are enthusiastic and eager to learn, and a group that is more passive and less enthusiastic. This obviously influences to outcome of the experiment because it is to be expected that the behavior of the "enthusiastic" group is different from the control group, and it cannot be said for certain anymore that this difference in behavior is explained by, and only by, the impact of the learning environment.

5.1.3. Sample size

In a perfect world it would be possible to calculate in advance the exact number of users that is needed to reach a statistically significant desired effect, for a certain metric. Even though we have a good idea of the expected mean and variance for each metric (since we have historical data at our disposal), it is hard to say in advance what effect size is realistic to expect. Consider the number of invoices created by new users with a known mean of 0.276 and a standard deviation of 1.18, over the first three weeks. What kind of increase in average number of invoices created is realistic to expect? This directly affects our sample size. Still, it is preferable to have a large sample size as this increases the minimum effect we can detect that is still statistically significant.

For practical reasons and because of time constraints we have chosen a three week period in which the test takes place. In the first two weeks users are selected for the A/B test given they fit the selection criteria. By selecting people for exactly two weeks we mostly even out any effect the day of the week may have on the sign up rate and user retention. Given the number of people signing up for Moneybird on an average day, and correcting for people failing to fit the selection criteria, we expect to have about 600 people in each group (the test group and control group) after two weeks. The users are then tracked for seven days each (see Figure 5.2). This means the first batch of users are part of the experiment from day 1 till day 7, and the last batch from day 14 till 20. We have chosen to track users for one week since we expect that the users can complete the lessons relatively quickly, and because we expect any possible effect as a result hereof to be visible in the first few days after completing the lessons.

5.1.4. Event tracking

For every major action the user may perform in the Moneybird application, an event is created. The event has a name (such as: `create_sales_invoice`), and some additional information about the event. To protect the privacy of the end user, the event contains a pseudo id, and not the actual id of the user. This way useful information and trends about groups of users can still be extracted, but the events by themselves are meaningless and cannot be traced back to individual users in the Moneybird application, as well as any personal information, or sensitive financial data they may have in their bookkeeping. In this experiment we utilize the event tracking system to detect differences in the group of users that is granted access to the learning environment, and the group that is not. We distinguish three categories of metrics:

1. Metrics related to bookkeeping reports available in Moneybird;
2. Metrics related to actionable bookkeeping features;
3. Metrics related to user engagement.

Bookkeeping report metrics

The selected bookkeeping metrics are summarized in Table 5.1. Via the main menu, the user can navigate to the reports page and this in turn triggers the creation of the `index_reports` event. The user can then navigate to one of the specific reports, such as the balance sheet and the profit loss statement, each of which create an event as well. Each report event includes the time period filter that the user applied to generate the report, but we will simply ignore this value and only look at the total event count instead. Other bookkeeping report metrics were ignored as well because they target specific reports that are available in Moneybird but were not addressed in the lessons of the learning environment.

Event	Description
`index_reports`	The user viewed the report overview page.
`balance_sheet_report`	The user viewed the balance sheet page.
`profit_loss_report`	The user viewed the profit loss statement page.

Table 5.1.: Overview of selected events related to the bookkeeping reports.

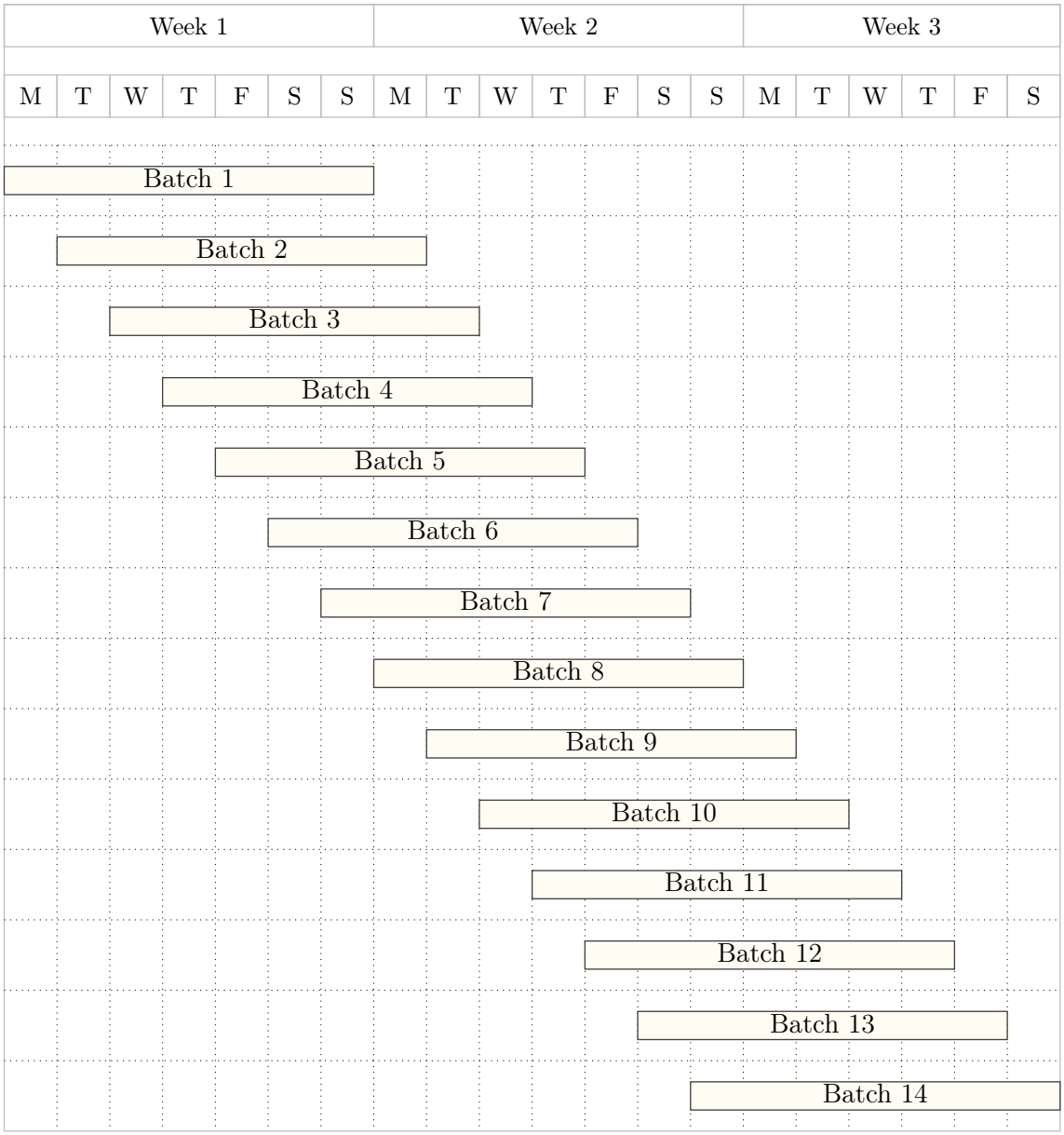


Figure 5.2.: Timeline of the A/B testing period. The test takes 20 days in total. Each batch represents a group of users that signed up on the given day and that fit the selection criteria described in section 5.1.1. Each user is tracked for seven days.

Bookkeeping action metrics

During the lessons in the learning environment, various actions are performed by the user, such as creating an invoice, linking transactions, and uploading a purchase invoice. We are interested in whether the users, as a result of these exercises, are more comfortable performing these same actions in Moneybird as well. Of course these actions also depend on the real life situation of the end user, since it makes sense the users only send an invoice if he or she actually has a client to bill. On the other hand, a invoice can be safely created and deleted before it is sent, and thus never reaching the client. The same goes for many other actions that more or less only affect the numbers on the reports, and can easily be reverted. Our assumption is that users might feel more comfortable experimenting with certain functionality in the main application as a result of completing the lessons in the learning environment. We simply measure this by comparing the sample mean of a metric between the test and control group, since it is hard to differentiate on the aforementioned behaviour. The actual number of purchases and sales within the business of the end users is of course not expected to differ as a result of being part of the test group or control group. Finally, users may carry over a bigger part (for example, further back in time) of their existing administration into Moneybird, as a result of partaking in the learning environment.

Just as in section 5.1.4, we made a selection of the events in the event tracking system that represent the actions we want to measure. In Table 5.2, the metrics are shown that represent key bookkeeping actions that we expect are performed more frequently by users in the test group.

Event	Description
`new_sales_invoice`	The user navigated to the new sales invoice page.
`create_sales_invoice`	The user created a sales invoice as draft.
`new_purchase_invoice`	The user navigated to the new purchase invoice page.
`create_purchase_invoice`	The user uploaded a purchase invoice, entered the invoice details and submitted the form.
`new_receipt`	The receipt form page was visited.
`create_receipt`	The user uploaded a receipt, entered the receipt details, and submitted the form.
`link_transaction`	A bank transaction was linked to a suggested category, sales invoice, or document.

Table 5.2.: Overview of selected events related to the bookkeeping actions.

User engagement metrics

With *user engagement* we denote the frequency with which the user interacts with the product. The term *adherence* is typically used to express the length of the time users stay active. Increased engagement and adherence indicates users that are more attached to the product, less likely to switch, and experience greater benefits from using it. Measuring user engagement is quite hard and might require a qualitative method instead, such as a survey or user interview, as opposed to the quantitative method (A/B test) used in this

research. In this study we make use of two meta tables in the event tracking system, that keep track of distinct hours and days during which a user was active. A user is defined as active if on a given day or hour at least one event was generated for that user in the event tracking system. Since a simple page visit in the application already generates an event, this is a precise enough definition of user activity for our experiment. Whether a user with a certain pseudo id is active on a moment can be looked up in the `hourly_active_users` and `daily_active_users` meta event tables, see Table 5.3. We use these tables to express user engagement as the total of distinct hours (or days) during which a user was active, across one week time. Finally, note that we are unable to measure *conversion* (the percentage of users that upgraded to a paid plan), since new users are allowed to use the product for free for two months, which means that during the A/B testing period they won't yet have to decide whether they want to upgrade or not. For this reason we simply excluded this metric.

Table	Description
`hourly_active_users`	Table with entries for each hour in which a user was active. Each row contains the timestamp of the hour and a pseudo id.
`daily_active_users`	Table with entries for each day on which a user was active. Each row contains the date and a pseudo id.

Table 5.3.: Meta events table with user activity per day and hour.

5.1.5. Statistical significance

We use a one-sided t-test to decide whether the difference in means of two Gaussian distributed test metric results is significant. The reason for using a one-sided test as opposed to a two-sided test is that we are not interested in whether a possible decrease is significant, only whether an increase is. If a decrease is found from the control group to the test group, we can safely conclude to reject the alternative hypothesis:

$$H_0 : \mu_{\text{control}} = \mu_{\text{test}}$$
$$H_a : \mu_{\text{test}} > \mu_{\text{control}}$$

where:

$$\mu_{\text{control}} = \text{the sample mean for metric } m \text{ across the control group.}$$
$$\mu_{\text{test}} = \text{the sample mean for metric } m \text{ across the test group.}$$

Note that the sample size of the control group and the test group need not be the same. A requirement for the t-test is that the sample means follow a Gaussian distribution and this is in fact the case for our test metrics. This may seem counterintuitive since the number of events generated by *a single user in a given time interval* may just be Poisson distributed. Even a Poisson distribution does not accurately reflect the behaviour of the users though, since not only the rate of events but also the possibility for a user to become inactive needs to be taken into account. A model that would approach the probability distribution of events generated by an individual user is the *Beta Geometric / Negative Binomial Distribution* (BG/NBD) model by Fader, Hardie and Lee [FHL05]. The sample mean of a given event across the test group could then be modeled as the sum of the BG/NBD probability distributions of the individual users. Finally, a statistical test could

be developed to decide whether the two probability distributions of both groups differ significantly.

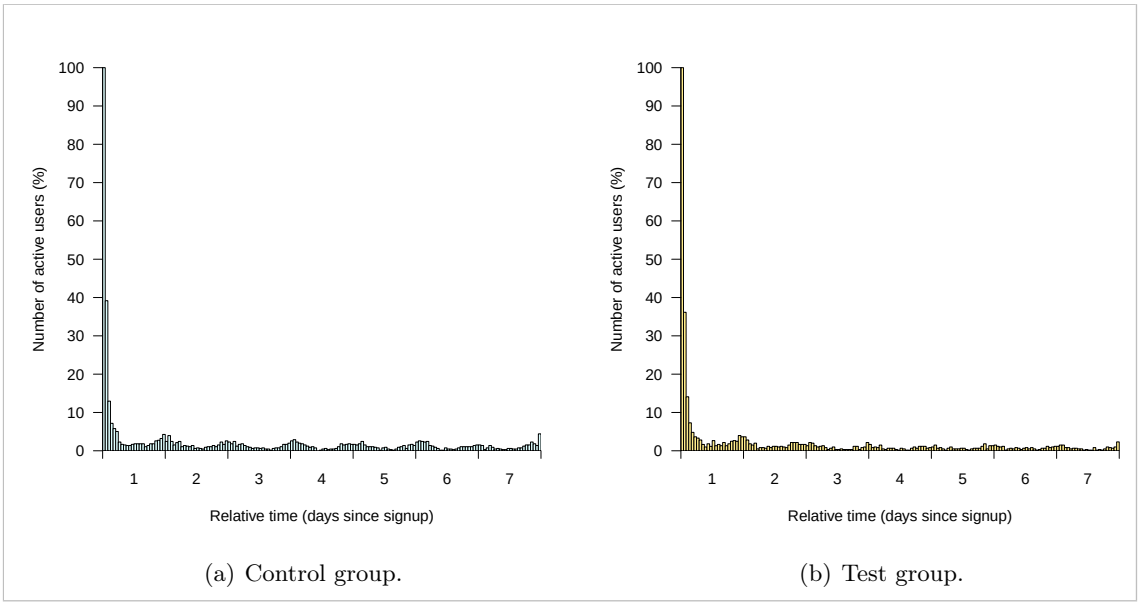
Luckily, we do not have to resort to such a complex model because in this case the sample size of both groups is sufficiently large that we can apply the Central Limit Theorem [Wik17b]. This theorem states that as the sample size increases, the sum (and thus the *sample mean*) approaches a normal distribution. What this means is that we can use the same approach (since all sample means approach a Gaussian distribution) for each proposed metric: compute the standard deviation and the difference in sample means between the test group and control group, apply a t-test with $\alpha = 0.05$, and decide whether the measured difference in sample means is statistically significant. To account for the expected difference in variance between both groups, we use Welch's t-test as opposed to the standard Student's t-test. Welch's t-test is known to be more robust and works better on skewed distributions and larger sample sizes [Zim04] [Fag12]. Finally, aside from the p-value, we also compute the effect size according to Cohen's d: the difference in sample means divided by the pooled standard deviation [Coh77]. The effect size makes it straightforward to see at a glance for which metric the largest effect was detected, and to see if a positive effect was found in the first place. Contrary to the p-value, the effect size does not take into account the sample size. This is fine when comparing our metrics since they are all based on the same sample size.

5.2. Results

For two weeks, numerous people signed up for a two-month free trial of Moneybird. In this section we review the results of the A/B test. These people were divided in two groups group, and tracked for one week each, as described in section 5.1.1. People that did not fit the criteria for either one of the groups were ignored. Since the only difference between the criteria of the test group and the control group is whether the pseudo id of a user is even or odd, it is to be expected that both groups are about the same size. After the experiment was over, **656 users** had been selected that fit the criteria of the control group, and **603 users** that fit the criteria of the test group. The difference can be explained by pure chance as it may happen that more users with an odd pseudo id to fit the selection criteria, than those with an even pseudo id. In the following sections we will see whether the behaviour of the test group was significantly different from the control group, according to our metrics discussed in section 5.1.4.

5.2.1. Overall activity

Before carrying out the statistical tests, and before analyzing the usage behaviour in the learning environment, let us first give an overview of the usage behaviour for users in the control and the test group in the main application. We found that in one week time (the time period during which users were tracked), a large percentage of the users were only active on the first day. The other users returned on at least one of the other six days in the first week. Of course, users that are only active on day one can also return in the following week, or month, but we did not investigate this statistic because we did not have the luxury of waiting for a prolonged period of time. A similar pattern shows when considering the percentage of people that were only active in the first hour after signing up. In Figure 5.3 and Table 5.4, the overall activity of users in the control and test group in this experiment is shown. Here we only considered what percentage of users were active on a given day or hour, not whether it was the only moment during which they were active in that week.



(a) Control group.

(b) Test group.

Figure 5.3.: Behaviour of users in the control and test group, per hour. A given user is active at a given hour if at least one event was generated for that user in that hour. In the first hour, 100% of users were active since that is when they signed up. Peaks are visible around each 24 hour interval.

Day (relative since signup)	1	2	3	4	5	6	7
Control group (active users)	100%	18%	14%	14%	12%	14%	13%
Test group (active users)	100%	18%	11%	9%	9%	9%	8%

Table 5.4.: Similar to Figure 5.3, for each day the relative number of active users is listed. On the first day all users are active since that is when they signed up. The total need not add up to 100%, because at any given day a varying percent of users can be active.

It becomes apparent that a fair bit of users are only active in the very beginning. This also somewhat shows from Table 5.5, which lists the number of users that used a certain feature. Sales invoices is by far the most popular feature. In the table, the usage of the learning environment is listed as well, for which we counted the number of users that completed at least one lesson in the first week. Of course it is not entirely fair to compare features in this way, because some features appeal to a broader audience, and other features are inherently more complex or time consuming to use (which does not necessarily make them any less valuable to the end user).

Feature	Metric	Control	Test
Sales invoices	$\#(sent\_sales\_invoice >= 1)$	146	121
Learning	$\#(lesson\_completed >= 1)$	-	37
Purchase invoices	$\#(create\_purchase\_invoice >= 1)$	19	19
Transactions	$\#(link\_transaction >= 1)$	20	16
Receipts	$\#(create\_receipt >= 1)$	10	7

Table 5.5.: Number of users per group that used a given feature at least once.

5.2.2. Learning environment usage

Of the 603 people in the test group, 114 accessed the learning environment. When splitting the test group into users that did not return to Moneybird after the first day they signed up, and users that did, we see that 12% of the non-returning users accessed the learning environment, versus 31% of the returning users. Let us dive in and see what the behaviour of these successfully logged in users looks like. In Figure 5.4, the number of lessons completed by the users in the learning environment is shown. Each lesson consists of multiple sections, for each of which a bar is shown in the histogram, with the number of people that completed that section.

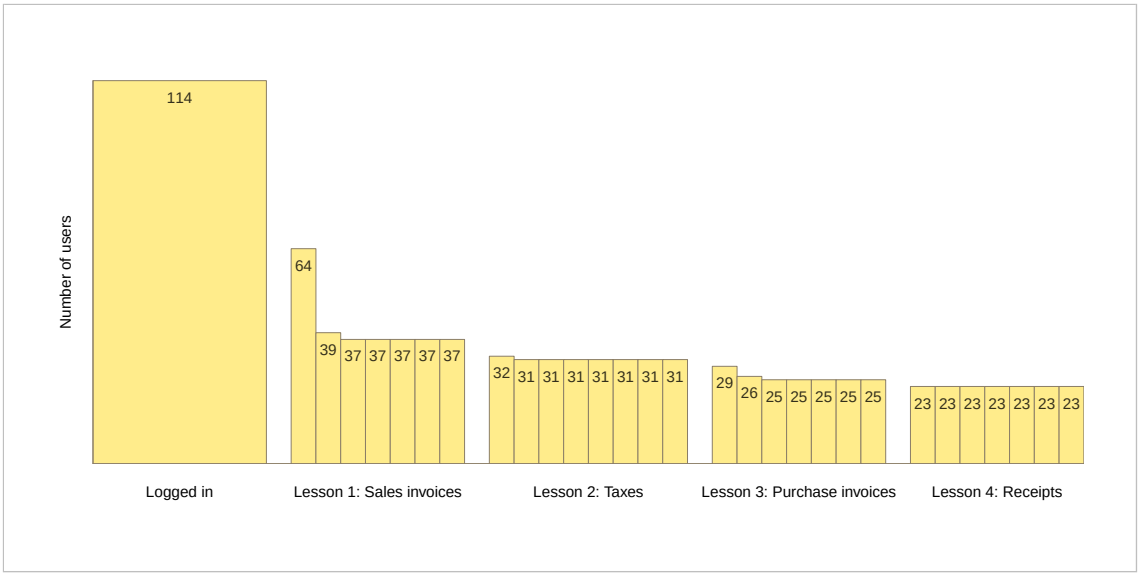


Figure 5.4.: Lessons and sections completed by the test group users that logged in to the learning environment. 37 users completed one lesson or more, and 23 finished all four lessons.

What stands out in Figure 5.4 is that a fairly large number of people showed interest in the learning environment, but most of these users did not complete any lessons. The second section of the first lesson is the first moment where users actually have to do something (complete an exercise), so this could explain the shape of the histogram. On the other hand, the users that did get past this point were fairly consistent in completing the other

lessons of the learning environment as well. In section 5.3, we discuss plausible explanations for this behaviour.

Since the learning environment was built to track and show the progress of the users throughout the lessons, we have access to the exact timestamps at which users completed a lesson or section. This allows us to estimate the time spent by users on each lesson. From the usability test that was conducted, we already had a rough idea (see Table 4.3). In Table 5.6, the time it took users on average to complete each lesson is shown. Only fully completed lessons were included. The time spent on a lesson was calculated as the sum of differences of the timestamps at which each of the subsequent sections in a lesson was completed. If a user took more than fifteen minutes to go from one section to the next, we assume they took a break. In those cases we estimated that the user spent five minutes in the learning environment and the rest of the time on their break.

Lesson	**Completion time** (`mm:ss`)
Lesson 1	5:15 (± 2:55)
Lesson 2	4:16 (± 2:19)
Lesson 3	5:09 (± 3:42)
Lesson 4	5:10 (± 3:53)

Table 5.6.: Average time spent to complete each lesson.

Let us conclude this section about the learning environment by summing up the performance on the multiple choice question. In Figure 5.5, the distribution of number of attempts needed to answer the multiple choice questions in lessons 1 and 2 is shown. In Appendix D the results for lessons 3 and 4 are shown as well. Although it is difficult to say something in general about the multiple choice questions (since they differ in the number of answer options), we see that for most questions, most users chose the correct answer in one try. The first question of lesson four appeared to be the most difficult one. This is the only question were four possible answers were available. The chart shows that many users had to guess the right answer, since the number of attempts is almost uniformly distributed. The data collected provides great insight in which questions should be made a little more difficult, and which ones are too easy. Question 3 of lesson 2 was answered correctly in one attempt by all 31 users, for example. The data collected tells us how many attempts users needed, but also which incorrect answer were picked the most, and which were picked the least. In order to increase the difficulty of a question, answers that are obviously incorrect can be interchanged with more plausible answers. Finally, it is important to find the right balance between accomplishment and challenge. If the questions are too easy, users may not learn anything new from them at all. On the other hand, if questions are too difficult, it is possible that users lose motivation, even though the mistakes made and the feedback shown after each attempt should ultimately lead to a better understanding of the lesson material.

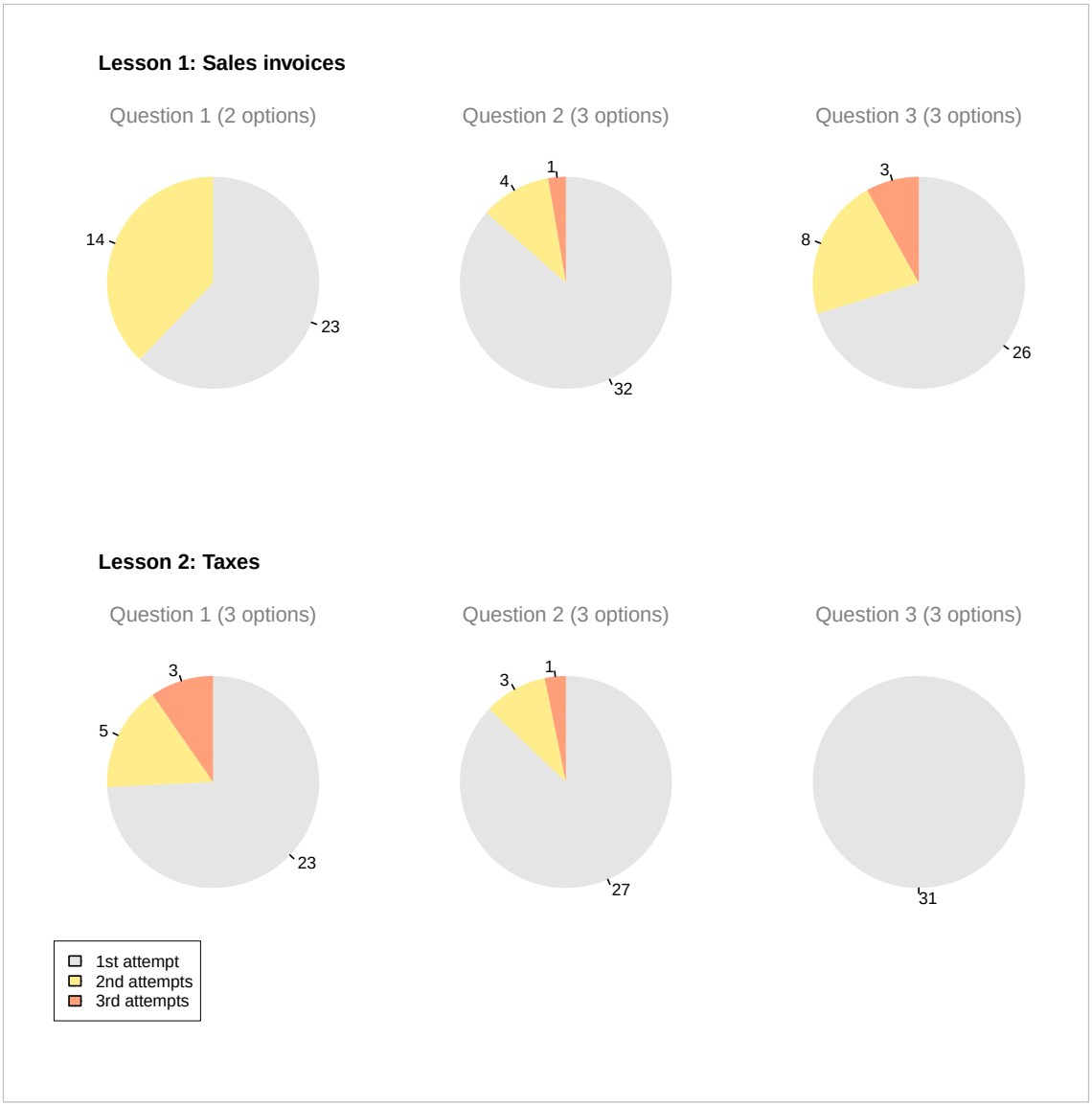


Figure 5.5.: The distribution of users that needed a given number of attempts needed to correctly answer the multiple choice questions. *Note that the first question of lesson 1 has two possible answers while the others have three.*

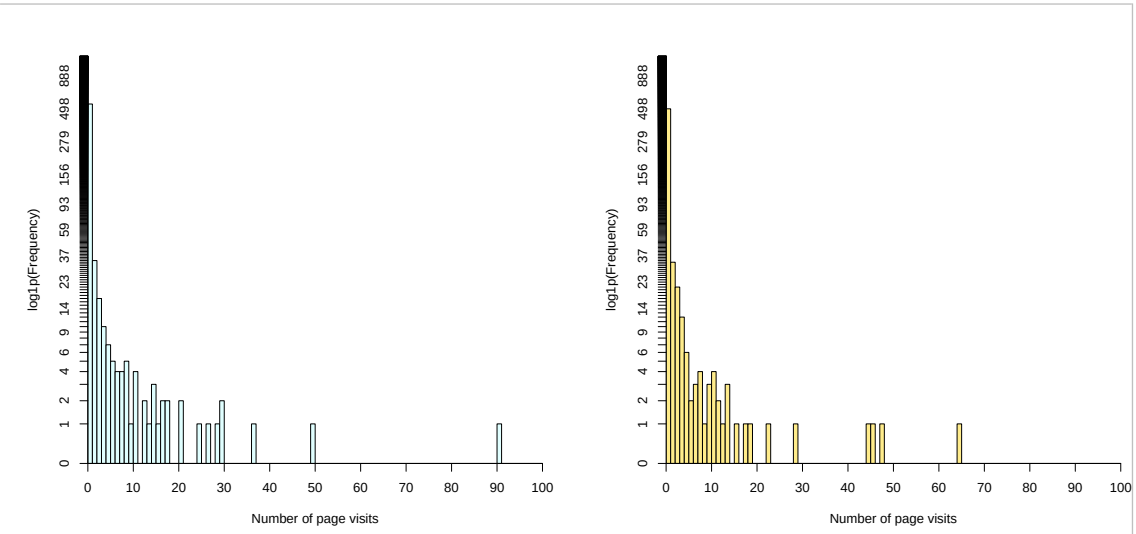
Now that we have discussed the usage behaviour in the learning environment, let us review final results of this research: whether any of our metrics showed a significant increase between the control group and the test group.

5.2.3. Bookkeeping activity results

In Table 5.7, the results for the metrics from section 5.1.4 and 5.1.4 are shown. Statistical significance was tested as described in section 5.1.5 and the resulting p-values and effect sizes are included as well. None of the p-values come close to the chosen significance level of $\alpha = 0.05$. The means μ and standard deviations σ per metric were calculated from the sums of events per user for that metric. In Figure 5.6, the frequency distribution of the `index_reports` event is shown. A logarithmic scale is used on the y-axis to better show low frequencies. The bulk of users did not visit the reports index page (74%). One fanatical user in the control group visited the page 408 times, though.

	μ_{control}	σ_{control}	μ_{test}	σ_{test}	Effect size	P-value
`index_reports`	[redacted]	9.69	[redacted]	5.00	-0.07	0.877
`balance_sheet_report`	[redacted]	2.59	[redacted]	2.89	+0.04	0.263
`profit_loss_report`	[redacted]	5.67	[redacted]	2.12	-0.04	0.736
`new_purchase_invoice`	[redacted]	17.4	[redacted]	4.57	-0.08	0.930
`create_purchase_invoice`	[redacted]	3.56	[redacted]	1.16	-0.04	0.738
`new_receipt`	[redacted]	2.05	[redacted]	1.27	-0.08	0.921
`create_receipt`	[redacted]	0.40	[redacted]	0.35	-0.04	0.733
`new_sales_invoice`	[redacted]	3.68	[redacted]	3.30	-0.05	0.829
`create_sales_invoice`	[redacted]	3.68	[redacted]	3.30	-0.05	0.829
`link_transaction`	[redacted]	13.12	[redacted]	4.57	-0.05	0.812

Table 5.7.: Bookkeeping activity results. The number of events generated for each metric were tracked, for both groups.



(a) Control group.

(b) Test group.

Figure 5.6.: Histogram of the total number of times users visited the reports index page in one week time. One user in the control group visited the page 408 times, this value is not shown. Note that a logarithmic scale is used along the y-axis to better show low frequencies.

5.2.4. User engagement results

Aside from the bookkeeping activity, we were interested in whether users became more active in the main application as a result of being able to access the learning environment. The time span in which the activity was tracked is one week, starting at the moment the user had signed up. In Table 5.8 the results are listed for the metrics discussed in section 5.1.4. The mean and standard deviation are calculated from the number of distinct days (`daily_active_users`) and hours (`hourly_active_users`) in which the users were active.

	μ_{control}	σ_{control}	μ_{test}	σ_{test}	Effect size	P-value
`hourly_active_users`	[redacted]	4.66	[redacted]	3.77	-0.09	0.9481
`daily_active_users`	[redacted]	1.31	[redacted]	1.16	-0.08	0.9295

Table 5.8.: User engagement results. The number of unique hours and days in which the user was active during the first week was measured.

5.3. Conclusion

Let us now sum up the results of the A/B test. We aimed to derive bookkeeping proficiency conclusions from a possible increase in activity in the main application, by users who had access to online learning environment. This was measured via numerous metrics which were collected via the existing event tracking architecture.

None of the proposed metrics reached statistical significance. We therefore unfortunately cannot reject the null hypothesis in favor of the alternative hypothesis, for any of the metrics. It seems that overall, users in the test group were even slightly less active (also when taking into account the smaller sample size). This shows from Table 5.4 as well as the effect sizes listed in 5.7 and 5.8. New users seem to have a fixed amount of time, and they choose to spend it on other activities, possibly on the learning environment instead. We saw that for most metrics, a slight negative effect size was detected. However, this difference was not found to be significant when performing an additional two-sided t-test. We further saw that the event data (when plotted in a histogram) is very skewed towards the left: a fair bit of users are only active right after they sign up, and then disappear (or at least not return in that same week). Moneybird offers a two-month free trial so in that regard it makes sense that many users only briefly explore the application, and then decide later on whether they want to use it for their full bookkeeping. Perhaps they are still comparing various bookkeeping software, or perhaps they are in the beginning phase of their business and simply do not yet have much bookkeeping to do. Of the 603 people that were included in the test group, 114 logged in to the learning environment. This includes both users that remained active in the main application throughout the first week, as well as users that disappeared after day one. Of these 114 users, we saw that 37 users completed one lesson or more. Beyond that point, we saw that many users made it all the way through till the very end, with a couple of users dropping out after lesson 1 and at the start of lesson 3 (see Figure 5.4).

What is perhaps most surprising is that of the 114 users that showed interest in the learning environment, only a small group completed any lessons. This seems to resemble the overall behaviour of new users in Moneybird. There are a few explanations that can be thought of:

Possible technical issues The application was fairly well tested and error logging was in place to detect any technical errors, but nothing major was found. It was known upfront, though, that about 7% of the users would not be able to use the learning environment because of their outdated browser version, and another 7% would have a slightly less optimal user experience. However, this by itself cannot entirely account for the number of users that dropped out. Furthermore, no support tickets were received regarding any issues that may have been encountered. Browser issues could have played a minor role, though.

Lessons may have been too easy Because of the focus on beginners, and because the lessons can only be completed in order (it is not possible the skip the first lesson, for example), it could be that many users decided they would not benefit from partaking in the learning environment. On the other hand, even if for some users the lessons were too easy, they might still want to complete them (out of curiosity or competitiveness). This seems to be the case as many users did quite well on the multiple choice questions. Still, a broader range in difficulty and the ability to start with any lesson at any point in time, would be desired to achieve a better retention rate.

The behaviour is inherent to that of newly signed up users Since there is such a resemblance to the behaviour of new users in Moneybird, it could be that the usage behaviour in the learning environment is simply a result of characteristics inherent to this group of users. Many users are only active on the first day, as if they are only exploring and do not intend to do any actual bookkeeping. After the initial spike, the remaining users stay about equally active from day to day and from lesson to lesson. Table 5.4 and Figure 5.4 nicely show this resemblance. Because of the setup of the experiment, we were unfortunately unable to ask users why they left. Further research is required to better understand their motives.

Although it is hard to say for certain, we think that drop out rate from lesson to lesson is mostly explained by the second reason (the lessons being too easy for some users), and that the initial drop out rate is mostly explained by the third reason (it is inherent to that of new users). This hypothesis is supported by the data shown in Table 5.5, from which we learned that (when compared to some other key features of Moneybird) the level of engagement in the learning environment is actually relatively high, and is only outperformed by the *sales invoices* feature in the main application.

Finally, even though only a small group completed any lessons at all, those that did, rarely dropped out in the middle of a lesson and most of them even completed all four lessons. It seems that for a specific audience the learning environment did really well. Going by the number of mistakes made in answering the multiple choice questions, some users may have desired something a bit more challenging, but overall the questions were not too hard and not too easy. If we look at the time spent by users on the lessons, we see that they spent about five minutes on each (although this did vary quite a bit per user). Some users even spent over an hour to complete all four lessons, possibly taking breaks in between to read and learn about bookkeeping via other resources, or simply because they got distracted. In the final chapter of this study we discuss the role of the e-learning and persuasive design techniques that were applied, the value of the usability test, the setup of the final experiment, and what could have been done differently.

6. Conclusion

Let us conclude this study by summing up the methodology used and main result achieved throughout this research, in order to answer the research questions posed back in section 1.3. A short discussion and corresponding recommendations for future work are provided in section 6.2 and 6.3, respectively.

6.1. Summary

Firstly, in Chapter 1, we explored the context in which the research takes places: an online bookkeeping platform with over 150.000 users. The motivation for improving the bookkeeping skills of these users quickly became clear as the benefits were analyzed: a potential decrease in support tickets and greater overall user satisfaction. To narrow down the target audience, we decided to focus on newly signed up users, because overall these users have the biggest potential to learn. We chose to teach these users the basics of bookkeeping through an online learning environment.

In Chapter 2, we learned about e-learning and persuasive design; the two research areas that are most relevant to developing the online learning environment. We saw that extensive research has been done in the field of e-learning. This previous work resulted in a set of design principles that we could directly apply in our current research. On the other hand, we explored persuasive design, a research area that focuses on supporting users in achieving a certain personal goal, through the use of technology. We learned that many opportunities exist to incorporate persuasive design into the learning environment, such as the use of points or badges as a reward system, and the use of a virtual coach to guide users through the lessons. The aspects of persuasive design that were incorporated in the learning environment have been discussed in section 3.4. Finally, we reviewed numerous existing online learning environments such as Khan Academy and Babbel, and tried to analyze what made them successful. The use of gamification was briefly discussed as a sub category of persuasive design.

In the next chapter, Chapter 3, the decisions made while designing the actual learning environment were explained, as well as how they connect to the reviewed literature in the previous chapter. The result is a web-based application that can be accessed by anyone with a Moneybird account and contains four lessons about the basics of online bookkeeping. A combination of interactive exercises, multiple choice questions with tailored feedback per answer, and illustrative graphics in the form of financial reports were used to communicate the learning material in an engaging and understandable way. The learning material itself was created in collaboration with two bookkeeping experts at Moneybird.

To evaluate the learning environment, a two-step approach was taken. First a usability test was conducted, as seen in Chapter 4. Six people new to Moneybird were invited to partake in one-hour sessions, during which the test facilitator observed the behaviour of these users as they navigated through the learning environment. An overview of problems encountered by the users was provided as a result hereof. No major issues were encountered by the users and the overall feedback was positive. The minor issues that were found mainly concerned a few parts of the application that were not entirely intuitive for the new users. These issues were fixed before the start of the next, and final experiment.

The final experiment that was carried out is described in Chapter 5. In this experiment,

the learning environment was subjected to a group of test users, and the behaviour of these test users in the main application was compared to that of a control group. The behaviour of users was anonymously tracked for one week using the existing event tracking architecture at Moneybird. For two weeks, people that signed up to Moneybird were randomly divided into two groups, filtered by predetermined selection criteria. This resulted in a control group of 656 users and a test group consisting of 603 users. The test users saw a slightly altered version of the main application that included links to the learning application. 114 test users logged in to the learning environment throughout the course of the experiment, 37 of these users completed one lesson or more. We then analyzed whether the behaviour of the test group in the main application differed significantly from the control group. We specifically looked at several metrics that were chosen beforehand, metrics that could indicate a better understanding of bookkeeping. We also included two metrics to express the overall level of activity of users in the main application. We saw a surprising result in that many of the proposed metrics showed a decrease in comparison to the control group, as opposed to an increase. Overall it seemed that the users of the control group were slightly more active. The difference was not significant though, according to the predetermined significance level of $\alpha = 0.05$. An optimistic explanation for these findings is that users in the test group spent more time in the learning environment and as a result hereof less time in the main application. However, we cannot say for certain whether this is the case based on the data that was collected in the relatively short amount of time. In this chapter we also learned about the large variance that exists in the behaviour of new users of Moneybird. For unknown reasons, many users are only active on the first day and do not return anymore afterwards. The ease of signing up and the existence of other competing free bookkeeping software could be one of the reasons. We saw a similar pattern in the learning environment, where many users were only initially active. Finally, another plausible explanation for the drop out rate is that for more advanced users the bookkeeping course is too easy. However, on the other hand we saw that most users that got past the initial hurdle stayed active throughout the rest of the course and completed all of the four available lessons.

6.2. Discussion

The methodology described in this research was carried out according to plan, but still along the way we gained new insights on which things could be done differently, and which opportunities for future work still exist. In this section a number of remarks are made regarding the methodology used in this research study. For each remark, a corresponding recommendation for future work is given in section 6.3.

Lesson material The lesson material that was created during this research consists of four topics about the basics of bookkeeping. The number of lessons available directly impacts the potential knowledge gain of users, and thus the outcome of our final experiment. Because of the chosen target audience of new users (presumably beginners) we decided to focus on the easy topics at first, and due to time constraints we were unable to provide a wider range of lessons. This is obviously something that could be addressed in future developments.

Final experiment Because of how the final experiment was set up, we did not have control over how many users would actually access the learning environment. The ratio of users that use the learning environment directly impacts the effect size that can be achieved, since it is these users that potentially show different behaviour in the main application as a result of the bookkeeping lessons they completed. We could only estimate in advance the sample size of the test group after a given period of

time, since we knew how many users sign up every day. Furthermore, the scope of the changes evaluated in the A/B test is fairly broad: a whole new application is made available. Regardless of the outcome of the test, the broad scope made it difficult to draw conclusions on the role of the specific e-learning and persuasive design principles that were implemented. Only a conclusion on the overall impact of the learning environment can be drawn, and the exact motivation of users remains unclear. This brings us to the next point.

Qualitative data Even though a lot of *quantitative* data was collected in the final experiment, we did not collect as much *qualitative* data. The learning environment did include the option to send feedback via email, but none of the users did. It would have been especially valuable to know more about the reasons users left the application. During the usability test we *did* collect qualitative data, although on a fairly small scale. Furthermore, the focus was on detecting usability problems and not so much on why users would or would not use the learning environment. This is also difficult to gauge during a usability test. The users that were invited to partake in the usability test would need to make a judgment call about how they would prioritize the learning environment in the future, when they are at home or at work, and many other things require their attention as well. Finally, with additional qualitative data, we would have been better able to understand why a fairly large number of users did not complete more than one lesson, and why the users in the test group seemed slightly less active in the main application, than those in the control group. In conclusion, many opportunities still exist to learn more about the motives of users in the learning environment, and Moneybird in general.

6.3. Recommendations

Finally, let us conclude this research study with a few recommendations for future work.

Lesson material First and foremost, our recommendation is to cover a broader range of bookkeeping topics. Especially the more advanced topics are valuable to cover because this increases the target audience that can be served. Not only beginners, but also users with preexisting bookkeeping knowledge can then learn from the lessons available. Furthermore, our recommendation is to increase the difficulty of the lessons by including another type of content: *challenges*. Challenges are an extension of exercises with the difference that in order to be completed they require a better understanding of the lesson material. An example of a challenge could be to complete the missing values on a profit loss report. It is then much harder to guess (as is possible in multiple choice question) and in general much more challenging than the exercises where the user simply has to follow the instructions. Finally, the addition of narrated videos would make the lesson material even more approachable. This idea is supported by the modality principle, as discussed in section 3.3.

Final experiment To better quantify the impact of each of the e-learning and persuasive design principles, our recommendation is to perform *multivariate testing* instead. With multivariate testing, multiple variations can be compared simultaneously. In each variation, only a single element is changed. Version *A* would still include all functionality, but additional variations would leave out the implementation of a single design principle. We can then track the impact of each change individually. With more variations, the sample size is larger, although on the other hand we only consider users that visited the learning environment in the first place, as opposed to all new users of Moneybird. Note that this approach no longer allows us to compare against

a group of users that did not use the learning environment, but it does address the issues described in section 6.2.

Qualitative data The way Moneybird is currently developed is based on an overall vision of what the company stands for, feedback received from end users via support tickets or social media, and by observing patters in user behaviour via data analysis. In this research, to answer the research questions, we also relied heavily on the collection of quantitative data. In future research, the focus should be on collecting additional qualitative data, on top of the feedback collected during the usability test. Feedback of users should be collected via user interviews, as well as via surveys that can be directly integrated in the learning environment. Questions that can be asked in the survey are whether the user felt like they learned something while completing the lessons in the learning environment, and what content they would like to see the most in the future. Actively seeking out users that left the application and conducting user interviews with these people is vital too, in order to obtain a complete view. An incentive may need to be provided for these users to participate in such an interview. Ultimately, with the collection of additional qualitative data, a better understanding of the observed data is achieved. We think that this approach is the missing link in gaining a better understanding of the effectiveness of the learning application in increasing the bookkeeping proficiency of end users of Moneybird.

Bibliography

[Atk02] Robert K Atkinson. Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, 94(2):416, 2002.

[Ayr06] Paul Ayres. Impact of reducing intrinsic cognitive load on learning in a mathematical domain. *Applied Cognitive Psychology*, 20(3):287–298, 2006.

[BKS10] Paul Blayney, Slava Kalyuga, and John Sweller. Interactions between the isolated–interactive elements effect and levels of learner expertise: Experimental evidence from an accountancy class. *Instructional Science*, 38(3):277–287, 2010.

[BMS+96] Isabel L Beck, Margaret G McKeown, Cheryl Sandora, Linda Kucan, and Jo Worthy. Questioning the author: A yearlong classroom implementation to engage students with text. *The Elementary School Journal*, 96(4):385–414, 1996.

[But06] Kirsten R Butcher. Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*, 98(1):182, 2006.

[CAS05] Tracey Clarke, Paul Ayres, and John Sweller. The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications. *Educational Technology Research and Development*, 53(3):15–24, 2005.

[CFC05] Fiona Concannon, Antoinette Flynn, and Mark Campbell. What campus-based students think about the quality and benefits of e-learning. *British journal of educational technology*, 36(3):501–512, 2005.

[Che13] Sheng-Lun Cheng. *The Design of Online Experiential Financial Literacy Games*. PhD thesis, New Mexico State University, 2013.

[CJDA02] Heidi S Chumley-Jones, Alison Dobbie, and Cynthia L Alford. Web-based learning: Sound educational method or hype? a review of the evaluation literature. *Academic medicine*, 77(10):S86–S93, 2002.

[CM08] Ruth Colvin Clark and Richard E Mayer. Learning by viewing versus learning by doing: Evidence-based guidelines for principled learning environments. *Performance Improvement*, 47(9):5–13, 2008.

[CM16] Ruth C Clark and Richard E Mayer. *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & Sons, 2016.

[Coh77] Jacob Cohen. Statistical power analysis for the behavioral sciences (revised ed.), 1977.

[CS91] Paul Chandler and John Sweller. Cognitive load theory and the format of instruction. *Cognition and instruction*, 8(4):293–332, 1991.

[CY10] Cheng-Chieh Chang and Fang-Ying Yang. Exploring the cognitive loads of high-school students as they learn concepts in web-based environments. *Computers & Education*, 55(2):673–680, 2010.

[DDKN11] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*, pages 9–15. ACM, 2011.

[DF06] Sara I De Freitas. Using games and simulations for supporting learning. *Learning, media and technology*, 31(4):343–358, 2006.

[DFH07] M. Hakel D. F. Halpern, A. Graesser. 25 learning principles to guide pedagogy and the design of learning environments. *Association of Psychological Science Taskforce on Lifelong Learning at Work and at Home*, 2007.

[DSM+13] Nor Aziah Daud, Noraidah Sahari, Zurina Muda, et al. An initial model of persuasive design in web based learning environment. *Procedia Technology*, 11:895–902, 2013.

[Ebn07] Martin Ebner. E-learning 2.0= e-learning 1.0+ web 2.0? In *Availability, Reliability and Security, 2007. ARES 2007. The Second International Conference on*, pages 1235–1239. IEEE, 2007.

[Fag12] Morten W Fagerland. t-tests, non-parametric tests, and large studies—a paradox of statistical practice? *BMC Medical Research Methodology*, 12(1):78, 2012.

[FHL05] Peter S Fader, Bruce GS Hardie, and Ka Lok Lee. "counting your customers" the easy way: An alternative to the pareto/nbd model. *Marketing science*, 24(2):275–284, 2005.

[Fog02] Brian J Fogg. Persuasive technology: using computers to change what we think and do. *Ubiquity*, 2002(December):5, 2002.

[Fog09] Brian J Fogg. A behavior model for persuasive design. In *Proceedings of the 4th international Conference on Persuasive Technology*, page 40. ACM, 2009.

[Gin05] Paul Ginns. Meta-analysis of the modality effect. *Learning and Instruction*, 15(4):313–331, 2005.

[GPM+03] Pascal WM Gerven, Fred Paas, Jeroen JG Merriënboer, Maaike Hendriks, and Henk G Schmidt. The efficiency of multimedia learning into old age. *British journal of educational psychology*, 73(4):489–505, 2003.

[HHH09] Jana Holsanova, Nils Holmberg, and Kenneth Holmqvist. Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology*, 23(9):1215–1226, 2009.

[HKP14] Juho Hamari, Jonna Koivisto, and Tuomas Pakkanen. Do persuasive technologies persuade?-a review of empirical studies. In *International Conference on Persuasive Technology*, pages 118–136. Springer, 2014.

[HM98] Shannon F Harp and Richard E Mayer. How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of educational psychology*, 90(3):414, 1998.

[HMS07] Egbert G Harskamp, Richard E Mayer, and Cor Suhre. Does the modality principle for multimedia learning apply to science classrooms? *Learning and Instruction*, 17(5):465–477, 2007.

[IMS+13] Nabil Issa, Richard E Mayer, Mary Schuller, Edward Wang, Michael B Shapiro, and Debra A DaRosa. Teaching for understanding in medical classrooms using multimedia design principles. *Medical education*, 47(4):388–396, 2013.

[Kar10] Günizi Kartal. Does language matter in multimedia learning? personalization principle revisited. *Journal of Educational Psychology*, 102(3):615, 2010.

[KCS00] Slava Kalyuga, Paul Chandler, and John Sweller. Incorporating learner experience into the design of multimedia instruction. *Journal of educational psychology*, 92(1):126, 2000.

[KKvM06] Liesbeth Kester, Paul A Kirschner, and Jeroen JG van Merriënboer. Just-in-time information presentation: Improving learning a troubleshooting skill. *Contemporary Educational Psychology*, 31(2):167–185, 2006.

[KZSK14] Aïmen Khacharem, Bachir Zoudji, Ingrid AE Spanjers, and Slava Kalyuga. Improving learning from animated soccer scenes: Evidence for the expertise reversal effect. *Computers in Human Behavior*, 35:339–349, 2014.

[Lie15] Andreas Lieberoth. Shallow gamification: Testing psychological effects of framing an activity as a game. *Games and Culture*, 10(3):229–248, 2015.

[LMMP+14] Derick Leony, Pedro J Muñoz-Merino, Abelardo Pardo, José A Ruipérez-Valiente, David Arellano Martín-Caro, and Carlos Delgado Kloos. Rule-based detection of emotions in the khan academy platform. *International Workshop on Massive Open Online Courses, At Antigua, Guatemala*, 2014.

[LSMH07] Stephen Lehman, Gregory Schraw, Matthew T McCrudden, and Kendall Hartley. Processing and recall of seductive details in scientific text. *Contemporary Educational Psychology*, 32(4):569–587, 2007.

[LZMM06] Andrés Lucero, Rodrigo Zuloaga, Selene Mota, and Felipe Muñoz. Persuasive technologies in education: improving motivation to read and write for children. In *International Conference on Persuasive Technology*, pages 142–153. Springer, 2006.

[MA91] Richard E Mayer and Richard B Anderson. Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of educational psychology*, 83(4):484, 1991.

[MA92] Richard E Mayer and Richard B Anderson. The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of educational Psychology*, 84(4):444, 1992.

[May89] Richard E Mayer. Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology*, 81(2):240–46, 1989.

[May05a] Richard E Mayer. *The Cambridge handbook of multimedia learning*. Cambridge university press, 2005.

[May05b] Richard E Mayer. Principles for managing essential processing in multimedia learning: Segmenting, pretraining, and modality principles. *The Cambridge handbook of multimedia learning*, pages 169–182, 2005.

[MC01] Richard E Mayer and Paul Chandler. When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of educational psychology*, 93(2):390, 2001.

[MD12] Richard E Mayer and C Scott DaPra. An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied*, 18(3):239, 2012.

[MDM03] Richard E Mayer, Gayle T Dow, and Sarah Mayer. Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based microworlds? *Journal of educational psychology*, 95(4):806, 2003.

[MDM11] Bruce M McLaren, Krista E DeLeeuw, and Richard E Mayer. A politeness effect in learning with web-based intelligent tutors. *International Journal of Human-Computer Studies*, 69(1):70–79, 2011.

[MFFC04] Richard E Mayer, Sherry Fennell, Lindsay Farmer, and Julie Campbell. A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology*, 96(2):389, 2004.

[MGJR08] Richard E Mayer, Emily Griffith, Ilana TN Jurkowitz, and Daniel Rothman. Increased interestingness of extraneous details in a multimedia science presentation leads to decreased learning. *Journal of Experimental Psychology: Applied*, 14(4):329, 2008.

[MGK+14] Robert Murphy, Larry Gallagher, Andrew E Krumm, Jessica Mislevy, and Amy Hafter. Research on the use of khan academy in schools: Research brief. 2014.

[MHL01] Richard E Mayer, Julie Heiser, and Steve Lonn. Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of educational psychology*, 93(1):187, 2001.

[ML07] Catherine McLoughlin and Mark JW Lee. Social software and participatory learning: Pedagogical choices with technology affordances in the web 2.0 era. In *ICT: Providing choices for learners and learning. Proceedings ascilite Singapore 2007*, pages 664–675, 2007.

[MLG14] Baptiste Monterrat, Elise Lavoué, and Sébastien George. Motivation for learning: Adaptive gamification for web-based learning environments. In *6th International Conference on Computer Supported Education (CSEDU 2014)*, pages 117–125, 2014.

[MLS95] Seyed Yaghoub Mousavi, Renae Low, and John Sweller. Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of educational psychology*, 87(2):319, 1995.

[MM98] Richard E Mayer and Roxana Moreno. A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of educational psychology*, 90(2):312, 1998.

[MM99] Roxana Moreno and Richard E Mayer. Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of educational psychology*, 91(2):358, 1999.

[MM00] Roxana Moreno and Richard E Mayer. A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational psychology*, 92(1):117, 2000.

[MM02] Roxana Moreno and Richard E Mayer. Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology*, 94(1):156, 2002.

[MMBV99] Richard E Mayer, Roxana Moreno, Michelle Boire, and Shannon Vagge. Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of educational psychology*, 91(4):638, 1999.

[MMP+12] Nicholas W Maynard, Preeti Mehta, Jonas Parker, Jeffrey Steinberg, et al. Can games build financial capability? financial entertainment: A research overview. Technical report, 2012.

[MMSL01] Roxana Moreno, Richard E Mayer, Hiller A Spires, and James C Lester. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and instruction*, 19(2):177–213, 2001.

[MMVK13] Pedro J Muñoz-Merino, José A Ruipérez Valiente, and Carlos Delgado Kloos. Inferring higher level learning information from low level data for the khan academy platform. In *Proceedings of the third international conference on learning analytics and knowledge*, pages 112–116. ACM, 2013.

[MMW02] Richard E Mayer, Amanda Mathias, and Karen Wetzell. Fostering understanding of multimedia messages through pre-training: Evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology: Applied*, 8(3):147, 2002.

[Mon17] Moneybird. Een memoriaalboeking maken. https://www.moneybird.nl/een-memoriaalboeking-maken/, 2017. [Online; accessed 24-Dec-2017].

[MRO10] Roxana Moreno, Martin Reislein, and Gamze Ozogul. Using virtual peers to guide visual attention during learning. *Journal of Media Psychology*, 2010.

[MS94] Richard E Mayer and Valerie K Sims. For whom is a picture worth a thousand words? extensions of a dual-coding theory of multimedia learning. *Journal of educational psychology*, 86(3):389, 1994.

[MSA+14] Ulrike IE Magner, Rolf Schwonke, Vincent Aleven, Octav Popescu, and Alexander Renkl. Triggering situational interest by decorative illustrations both fosters and hinders learning in computer-based learning environments. *Learning and Instruction*, 29:141–152, 2014.

[MSBM95] Richard E Mayer, Kathryn Steinhoff, Gregory Bower, and Rebecca Mars. A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development*, 43(1):31–41, 1995.

[Mun11] Cristina Ioana Muntean. Raising engagement in e-learning through gamification. In *Proc. 6th International Conference on Virtual Learning ICVL*, pages 323–329, 2011.

[oHS17] U.S. Dept. of Health and Human Services. Reporting usability test results. https://www.usability.gov/how-to-and-tools/methods/reporting-usability-test-results.html, 2017. [Online; accessed 16-September-2017].

[OKH09] Harri Oinas-Kukkonen and Marja Harjumaa. Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems*, 24(1):28, 2009.

[PBB+07] Harold Pashler, Patrice M Bain, Brian A Bottge, Arthur Graesser, Kenneth Koedinger, Mark McDaniel, and Janet Metcalfe. Organizing instruction and study to improve student learning: Ies practice guide. *National Center for Education Research*, 2007.

[PCS02] Edwina Pollock, Paul Chandler, and John Sweller. Assimilating complex information. *Learning and instruction*, 12(1):61–86, 2002.

[PMSB11] Babette Park, Roxana Moreno, Tina Seufert, and Roland Brünken. Does cognitive load moderate the seductive details effect? a multimedia study. *Computers in Human Behavior*, 27(1):5–10, 2011.

[RVLK15] Pedro J Ruipérez-Valiente, José A andc, Derick Leony, and Carlos Delgado Kloos. Alas-ka: A learning analytics extension for better understanding the learning process in the khan academy platform. *Computers in Human Behavior*, 47:139–148, 2015.

[RWT05] David L Rodgers and Beverly J Withrow-Thorton. The effect of instructional media on learner motivation. *International Journal of Instructional Media*, 32(4):333, 2005.

[SB15] Kathy Sierra and Bert Bates. *Badass: Making Users Awesome*. "O'Reilly Media, Inc.", 2015.

[Set16] Burr Settles. How we learn how you learn. http://making.duolingo.com/how-we-learn-how-you-learn, 2016. [Online; accessed 20-May-2017].

[SL10] Megan P Stevenson and Min Liu. Learning a language with web 2.0: Exploring the use of social networking features of foreign language learning websites. *CALICO journal*, 27(2):233–259, 2010.

[SM16] Burr Settles and Brendan Meeder. A trainable spaced repetition model for language learning. In *Proceedings of the Association of Computational Linguistics (ACL)*, pages 1848–1858, 2016.

[SSG+14] Katharina Scheiter, Anne Schüler, Peter Gerjets, Thomas Huk, and Friedrich W Hesse. Extending multimedia research: How do prerequisite knowledge and reading comprehension affect learning from text and pictures. *Computers in Human Behavior*, 31:73–84, 2014.

[Str12] Valerie Strauss. Does the khan academy know how to teach? https://www.washingtonpost.com/blogs/answer-sheet/post/how-well-does-khan-academy-teach/2012/07/27/gJQA9bWEAX_blog.html?utm_term=.e422b80642cb, 2012. [Online; accessed 09-May-2017].

[SWVGVM11] Ingrid AE Spanjers, Pieter Wouters, Tamara Van Gog, and Jeroen JG Van Merrienboer. An expertise reversal effect of segmentation in learning from animated worked-out examples. *Computers in Human Behavior*, 27(1):46–52, 2011.

[Ush05] Eiko Ushida. The role of students' attitudes and motivation in second language learning in online language courses. *CALICO journal*, pages 49–78, 2005.

[Ves09] Roumen Vesselinov. Measuring the effectiveness of rosetta stone. *Rosetta Stone*, 2009.

[VG12] Roumen Vesselinov and John Grego. Duolingo effectiveness study. *City University of New York, USA*, 2012.

[Vic12] Bret Victor. Learnable programming: Designing a programming system for understanding programs. http://worrydream.com/LearnableProgramming/, 2012. [Online; accessed 09-May-2017].

[Wik17a] Wikipedia. E-learning (theory) — Wikipedia, the free encyclopedia, 2017. [Online; accessed 11-May-2017].

[Wik17b] Wikipedia. E-learning (theory) — Wikipedia, the free encyclopedia, 2017. [Online; accessed 26-Nov-2017].

[WJM+08] Ning Wang, W Lewis Johnson, Richard E Mayer, Paola Rizzo, Erin Shaw, and Heather Collins. The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies*, 66(2):98–112, 2008.

[WM08] Leah E Wickersham and Patricia McGee. Perceptions of satisfaction and deeper learning in an online course. *Quarterly Review of Distance Education*, 9(1):73–83, 2008.

[Woo17] Ed Wood. With this app made by 100+ language experts, you'll be speaking a new language in 3 weeks. https://www.babbel.com/en/magazine/top-app-for-learning-languages-by-100-linguists, 2017. [Online; accessed 21-May-2017].

[Zim04] Donald W Zimmerman. A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology*, 57(1):173–181, 2004.

A. Recruitment survey

Printed version of the Google Forms survey that was set up for people interested in partaking in the usability study.

Deelname gebruikersonderzoek

Volgende week vrijdag gaan we de gebruiksvriendelijkheid testen van een nieuwe applicatie die je helpt leren boekhouden. Hiervoor hebben we natuurlijk testpersonen nodig! Wil je meedoen? Vul dan deze korte vragenlijst in.

Geselecteerde deelnemers die meedoen aan het volledige onderzoek op locatie ontvangen een Bol.com waardebon van €40.

Datum: vrijdag 25 augustus
Duur: 60 minuten
Locatie: Enschede

* Required

1. **Ben je bereid naar Enschede te komen voor een gebruikersonderzoek op vrijdag 25 augustus?** *
   Mark only one oval.

   - ◯ Ja
   - ◯ Nee *Skip to "Helaas!."*

Beschikbaarheid

2. **Op welk tijdstip zou je langs kunnen komen?** *
   Je kunt meerdere opties aanvinken, we kiezen dan een geschikt moment. De sessie zal niet langer dan één uur duren.
   Check all that apply.

   - ☐ In Enschede bij Moneybird op 25 augustus tussen 10:30u en 12:30u
   - ☐ In Enschede bij Moneybird op 25 augustus tussen 13:30u en 16:30u
   - ☐ Other: _____

Boekhoudervaring

3. **Hoe schat je je inhoudelijke kennisniveau in m.b.t. boekhouden?**
   Mark only one oval.

   - ◯ Beginner
   - ◯ Gemiddeld
   - ◯ Gevorderd
   - ◯ Expert

4. **Hoe lang maak je al gebruik van Moneybird?**
Mark only one oval.

- ◯ < 1 maand
- ◯ > 1 maand en < 3 maanden
- ◯ > 3 maanden en < 6 maanden
- ◯ 6 maanden - 1 jaar
- ◯ 1 jaar of langer
- ◯ Ik maak nog geen gebruik van Moneybird

5. **Welke onderdelen van jouw administratie houd je bij in Moneybird?** *
Check all that apply.

- ☐ Verkoopfacturen
- ☐ Inkomende documenten (facturen, bonnetjes, documenten)
- ☐ Offertes
- ☐ Banktransacties
- ☐ Other: _____

Vertel ons iets over jezelf

6. **Wat is je geslacht?** *
Mark only one oval.

- ◯ Man
- ◯ Vrouw

7. **Wat is je leeftijd?** *
Mark only one oval.

- ◯ Jonger dan 25
- ◯ Tussen 25 en 34
- ◯ Tussen 35 en 44
- ◯ Tussen 45 en 55
- ◯ Ouder dan 55

8. **Wat is de hoogste opleiding die je gevolgd hebt?** *
Mark only one oval.

- ◯ Basisschool
- ◯ VMBO (of vergelijkbaar)
- ◯ MBO (of vergelijkbaar)
- ◯ HBO
- ◯ WO
- ◯ Other: _____

9. **Welk beroep beoefen je?** *

_____

Skip to question 10.

Helaas!

We voeren dit onderzoek uit op 25 augustus op locatie in Enschede. Als je verhinderd of niet bereid bent om naar Enschede te komen, kun je helaas niet mee doen.

Stop filling out this form.

Contactgegevens

10. **Wat is je naam?** *

11. **Wat is je e-mailadres?** *

12. **Wat is je telefoonnummer?**

B. Informed consent form

The informed consent form that was signed by the participants in advance of the usability test. By signing this form the participants agreed that the session was recorded.

DEELNEMERSOVEREENKOMST

Learning @ Moneybird

DOEL

Het doel van het onderzoek is om meer te leren over de gebruiksvriendelijkheid, en mogelijkheden tot verbetering, van de ontwikkelde applicatie.

OMSCHRIJVING VAN DE PROCEDURE

De deelnemer wordt gevraagd om twee taken uit te voeren. Het is de bedoeling dat de deelnemer feedback geeft op datgene wat er gebeurt op het scherm en zijn/haar gedachtengang tijdens het doorlopen van de test.

DUUR

De sessie zal ongeveer een half uur, en maximaal drie kwartier duren. Het is mogelijk om tussendoor een korte pauze te nemen.

RISICO'S

De sessie wordt opgenomen zodat de feedback optimaal verwerkt kan worden. De opgenomen sessies worden niet verder verspreid en alleen intern bewaart zo lang dat nodig is voor het onderzoek.

Mocht de deelnemer vermoeid of geïrriteerd raken tijdens het doorlopen van de test dan is er de mogelijkheid om een pauze te nemen. De deelnemer mag ook beslissen om te stoppen. Dit kan op elk moment. In alle gevallen ontvangt de deelnemer de beloning.

VOORDELEN

Het deelnemen aan dit onderzoek vormt een hele waardevolle bijdrage aan de verdere ontwikkeling van het project. De deelnemer wordt beloond met een cadeaubon van Bol.com. De cadeaubon wordt per email verzonden na afloop van de deelname.

VERTROUWELIJKHEID

Er wordt zorgvuldig gehandeld omtrent de privacy van de deelnemers. Er worden geen herleidbare persoonlijke gegevens opgenomen in de uiteindelijke onderzoeksresultaten.

RECHTEN VAN DE DEELNEMER

- Deelname is vrijwillig

- De deelnemer mag op elk moment kiezen om te stoppen
- De deelnemer heeft het recht op de hoogte te worden gesteld van eventuele nieuwe informatie die van invloed is op hun deelname in het onderzoek

CONTACTINFORMATIE

- Thomas Brus <support@moneybird.com>

HANDTEKENING

- [X] Ik heb de bovenstaande voorwaarde gelezen en begrepen

25-08-2017

_____

Datum

_____ _____

Deelnemer Onderzoeker

C. Usability test reports

Usability reports produced as a result of the usability study that was conducted. The format is based on a template published on Usability.gov [oHS17].

(1) Answer to open-ended question by test facilitator	
Participant	T1
Page	Lesoverzicht
Description	
Q: Geef aan wat je eerste indruk is, wat je ziet op het eerste scherm, en wat je opvalt"	
A: "Valt me op dat het heel netjes is. Je weet direct waar je moet klikken".	

(2) Comment / recommendation by participant	
Participant	T1
Page	Lesoverzicht → 1. Je allereerste factuur → 1. Introductie
Description	
"Ik vind het rood een beetje... errr"	
Severity	
⊙ *Minor* ○ Serious ○ Critical	

(3) Comment / recommendation by participant	
Participant	T1
Page	Lesoverzicht → 1. Je allereerste factuur → 1. Introductie
Description	
Bij allereerste scherm (Introductie): "Hij geeft heel duidelijk aan wat je gaat leren"	

(4) Comment / recommendation by participant

Participant T2

Page Lesoverzicht → 1. Je allereerste factuur → 1. Introductie

Description

"Wat ik mij wel af vraag is, hoeveel tijd ik hier aan kwijt ben. Dat ik bijvoorbeeld zie: deze les neemt gemiddeld 10 minuten in beslag."

(5) Problem encountered

Participant T1

Page Lesoverzicht → 1. Je allereerste factuur → 2. Opdracht

Description

Bij opdracht 2, na verzenden factuur (door op enter te drukken): "Maar dan zie ik wel dat die direct verzonden is, dus begin ik me zorgen te maken."

Q: "Oke. Waarom maak je je dan zorgen?"

A: "Omdat ik het nog een keer zou willen controleren of het goed is gegaan of niet."

Q: "Wat zou je dan precies willen controleren?"

A: "Of ik alles goed heb ingevuld. En als ik op enter druk dan verwacht ik dat ik uit het vakje ga en daarna nog de mogelijkheid krijg om te controleren of alles goed is"

Q: "Hoe zou je dat voor je zien?"

A: "Nou bijvoorbeeld, er komt dan een ander scherm met 'dit is uw factuur, weet u zeker dat ... Kloppen de gegevens?' en dan onderaan een keuze 'ja' of 'nee': ik wil toch nog wat aanpassen'.

Notitie onderzoeker: Later bleek dat ze gewend is haar boekhouding in Excel te doen waar <enter> een andere functie heeft (verlaat input en ga naar volgende regel)

Severity

⊙ *Minor* ○ Serious ○ Critical

(6) Comment / recommendation by participant

Participant T2

Page Lesoverzicht → 1. Je allereerste factuur → 2. Opdracht

Description

"Heel duidelijk dat ik bij opdracht 2 ben".

(7) Problem encountered

Participant T2

Page Lesoverzicht → 1. Je allereerste factuur → 2. Opdracht

Description

Gebruiker verwacht onder de factuur een button met "OK" of "Verzend"

Severity

⊙ *Minor* ○ Serious ○ Critical

(8) Problem encountered

Participant T2

Page Lesoverzicht → 1. Je allereerste factuur → 2. Opdracht

Description

Na klikken op verzendknop. "Oh. Ah. Ja. Dat was dus een conceptfactuur. De factuur is verzonden. Maar zo verzend je toch geen factuur?"

Q: "Hoe bedoel je dat?"

A: "De btw is niet gespecificeerd. En uhmmm. Oh. Dat had ik nog niet gezien dat je naar beneden kon scrollen. En verder: de kleine lettertjes op een factuur, betalingstermijn, factuurnummer. Dit ziet er voor mij niet uit als een volledige factuur."

Severity

○ Minor ⊙ *Serious* ○ Critical

(9) Comment / recommendation by participant

Participant T3

Page Lesoverzicht → 1. Je allereerste factuur → 2. Opdracht

Description

Participant is not entirely sure whether the explanation text below the invoice exercise was revealed through the application scrolling down automatically, or by manually scrolling down by the participant.

Severity

⊙ *Minor* ○ Serious ○ Critical

(10) Comment / recommendation by participant

Participant T4

Page Lesoverzicht → 1. Je allereerste factuur → 2. Opdracht

Description

"Fijn dat deze indicatie (wijst naar sidebar) hier staat"

Q: "Waarom vind je dat fijn?"

A: "Dat geeft je een beetje een indicatie hoe lang het gaat duren, (en) waar je bent".

(11) Problem encountered

Participant T4

Page Lesoverzicht → 1. Je allereerste factuur → 2. Opdracht

Description

Participant indicates that he did not expect an explanation to be present below the invoice exercise.

Severity

⊙ *Minor* ○ Serious ○ Critical

(12) Problem encountered

Participant T5

Page Lesoverzicht → 1. Je allereerste factuur → 2. Opdracht

Description

Participant clicks on the invoice status label ("Draft") instead of the send button.

Severity

⊙ *Minor* ○ Serious ○ Critical

(13) Problem encountered

Participant T5

Page Lesoverzicht → 1. Je allereerste factuur → 2. Opdracht

Description

Participant notes that is visually unclear that it is possible to scroll down, after completing the invoice exercise.

Severity

⊙ *Minor* ○ Serious ○ Critical

(14) Comment / recommendation by participant

Participant T1

Page Lesoverzicht → 1. Je allereerste factuur → 3. Opdracht

Description

"Oh dus als ik het goed begrijp dan heeft dat programma, Moneybird, al gezien dat de klant betaald heeft en moet ik dat alleen nog koppelen aan de factuur (...), denk ik"

(15) Comment / recommendation by participant

Participant T1

Page Lesoverzicht → 1. Je allereerste factuur → 3. Opdracht

Description

"Ik vind het wél leuk dat je die groen balkjes (ziet / te zien krijgt)!"

(16) Comment / recommendation by participant

Participant T1

Page Lesoverzicht → 1. Je allereerste factuur → 3. Opdracht

Description

"Het is ook handig dat je de "wist je dat?" er tussen hebt staan, dat, als je iets mist, dat je dan weet waar je het zou kunnen vinden."

Q: "Hoe bedoel je dat precies?"

A: "Nou zeker voor iemand die net begint met boekhouden, dan mist die ineens een bedrag, maar als er dan staat "een nog-niet gekoppeld bedrag verschijnt aan de creditkant onder: te rubriceren transacties", dan weet je in ieder geval in ieder geval waar je moet zoeken om het te vinden."

(17) Problem encountered

Participant T4

Page Lesoverzicht → 1. Je allereerste factuur → 3. Opdracht

Description

After succesfully linking the bank transaction, the participant wonders whether it is possible (by clicking on the blue link icon) to change what he just did in case he had made a mistake.

Severity

⊙ *Minor* ○ Serious ○ Critical

(18) Problem encountered

Participant T6

Page Lesoverzicht → 1. Je allereerste factuur → 3. Opdracht

Description

Participant wonders whether the drop down option represents a bank transaction (initial thought) or the line above. After thinking aloud the participant comes to the correct conclusion.

Severity

⊙ *Minor* ○ Serious ○ Critical

(19) Observation about pathways participants took

Participant T1

Page Lesoverzicht → 1. Je allereerste factuur → 4. Meerkeuzevraag

Description

Participant asks whether it is possible to navigate back to the previous section. The participant moves the mouse over to the sidebar and hovers over the correct item. He answers "Yes" to his own question. The participant clicks on the item and the previous section appears on screen.

(20) Observation about pathways participants took

Participant T1

Page Lesoverzicht → 1. Je allereerste factuur → 4. Meerkeuzevraag

Description

Bij meerkeuze vraag: gebruiker gaat uit zichzelf terug om de toelichting van de vorige sectie opnieuw te lezen.

(21) Comment / recommendation by participant

Participant T1

Page Lesoverzicht → 1. Je allereerste factuur → 4. Meerkeuzevraag

Description

Na zien dat het antwoord goed is van de eerste meerkeuzevraag: "Yaay! ... (gelach) ... En ik vind de smiley leuk, ik ben gek op smileys"

(22) Problem encountered

Participant T3

Page Lesoverzicht → 1. Je allereerste factuur → 4. Meerkeuzevraag

Description

A bug is encountered: somehow the answer to the multiple choice question was already filled out. Resetting via "Account > Voortgang resetten" did not resolve the problem. After fully resetting the environment via the commandline, the problem was resolved.

Severity

○ Minor ○ Serious ⊙ *Critical*

(23) Answer to open-ended question by test facilitator

Participant T1

Page Lesoverzicht → 1. Je allereerste factuur → 5. Meerkeuzevraag

Description

Na meteen goed beantwoorden van de tweede meerkeuzevraag:

Q: "Stel je zou een ander antwoord hebben gekozen, bijvoorbeeld die daarboven, en je zou dan op de knop daaronder drukken, zou je dan een idee hebben wat er dan zou gebeuren?"

A: "Uhmmm... Dan hóóp ik, dat ik uitleg krijg, waarom die niet goed is"

Q: "Oke. Wat voor soort uitleg? Inhoudelijke uitleg? Of ...?"

A: "Ja inhoudelijke uitleg of, uhmm... misschien nog een keer verduidelijking van de begrippen? Zeker als beginner denk ik dat je crediteuren en debiteuren snel door elkaar kunt halen."

(24) Answer to open-ended question by test facilitator	
Participant	T4
Page	Lesoverzicht → 1. Je allereerste factuur → 5. Meerkeuzevraag

Description

Q: "Stel je voor dat je nu een ander antwoord had gegeven, bijvoorbeeld het bovenste antwoord. Wat zou je dan verwachten dat er dan gebeurt?"

A: "Dat er een uitleg bij staat waarom dat niet goed is. En wat het goede antwoord dan wel is."

Q: "Alternatief zou nog zijn dat je dan zelf het goede antwoord moet selecteren."

A: "Ja. Maar misschien wordt het dan meer gokken. (...)"

(25) Answer to open-ended question by test facilitator	
Participant	T6
Page	Lesoverzicht → 1. Je allereerste factuur → 5. Meerkeuzevraag

Description

Q: "Stel, bij de vorige vraag, als je het andere antwoord had aagegeven, bijvoorbeeld de bovenste, wat zou je dan verwachten dat er zou gebeuren?"

A: "Een rood balkje. En een uitleg wat het goeie antwoord wel had moeten zijn."

(26) Problem encountered

Participant T1

Page Lesoverzicht → 1. Je allereerste factuur → 6. Meerkeuzevraag

Description

De gebruiker vult bij sectie 6 opzettelijk een verkeerd meerkeuzeantwoord in uit nieuwsgierigheid. De interface geeft aan dat het antwoord fout is, plus toelichting. Er volgt een stilte en lichte verwarring.

"Oh moet ik nou uh... Ik weet nu niet wat ik moet doen"

Opmerking onderzoeker: het word in eerste instantie niet helemaal duidelijk waar de verwarring zit: lesinhoudelijk of UX. Er volgt een korte monoloog van de deelnemer over wat het goede antwoord zou moeten zijn.

"Ik kan niet verder, dus dan denk ik dan klik ik dat maar aan als het goede antwoord. En dat kan wel!"

Q: "Hoe komt het dat je dacht dat je niet verder kon? Terwijl, nu ben je natuurlijk verder."

A: "Omdat er na de opmerking niet staat van uhm... Vul alsnog een ander antwoord in. En je kunt verder nergens op klikken, het ziet eruit alsof je verder nergens op kunt klikken."

En ik kon niet verder hier op (wijst button onder meerkeuzevraag aan) klikken, dus denk ik misschien moet ik dan maar een ander antwoord geven en kijken wat er gebeurd."

Severity

○ Minor ⊙ *Serious* ○ Critical

(27) Observation about pathways participants took

Participant T3

Page Lesoverzicht → 1. Je allereerste factuur → 6. Meerkeuzevraag

Description

After answering the multiple choice question correctly, the participant skips reading the explanation that is given, and immediately continues to the next, and final, section of the lesson.

Severity

⊙ *Minor* ○ Serious ○ Critical

(28) Problem encountered

Participant T4

Page Lesoverzicht → 1. Je allereerste factuur → 6. Meerkeuzevraag

Description

Participant is slightly confused about the terminology. It is unclear that 'debitors' refers to the category on the balance sheet and not actual debitors (persons). Slight confusion exists about the plural use of debitors vs debitor, since in this lesson there is only one client and not multiple.

Severity

⊙ *Minor* ○ Serious ○ Critical

(29) Comment / recommendation by participant

Participant	T1
Page	Lesoverzicht → 1. Je allereerste factuur → 7. Samenvatting

Description

"Dat vind ik verwarrend"

Q: "Wat vind je verwarrend?"

A: "De winst is geboekt bij het versturen van de factuur. Maar als de factuur is verstuurd, dan heb ik nog niks. Voor hetzelfde geld betaald de klant niet. En dan heb ik geen winst" (...) "Ik zou verwachten dat er iets staat zoals 'de *verwachte* winst is geboekt".

"Voor de rest, hardstikke duidelijk".

Severity

○ Minor ⊙ *Serious* ○ Critical

(30) Comment / recommendation by participant

Participant T1

Page Lesoverzicht → 1. Je allereerste factuur → 7. Samenvatting

Description

"Ik had nog verwacht dat het rode blok groen zou worden."

Q: "Waarom had je dat verwacht?"

A: Nou omdat zegmaar dat rood is zegmaar dat je denkt: ik had het nog niet af. Dat is logisch. Maar dan denk je toch op een gegeven moment ik heb 'm af, dat er dan iets gebeurt, met dat rode blok. Want dat rode blok is best wel, uh... aanwezig. Dat leidt een beetje af"

"Terwijl de rest van de kleuren, dat hebben jullie (het) echt heel rustig, en duidelijk gehouden, zoals het blauw en het groen, en ja het rood steekt dan heel erg af er tegen."

Severity

⊙ *Minor* ○ Serious ○ Critical

(31) Comment / recommendation by participant

Participant T2

Page Lesoverzicht → 1. Je allereerste factuur → 7. Samenvatting

Description

"Ja. Dit is mij helemaal duidelijk. Zelfs mij!"

...

"(geluidje) Yaay! (deelnemer klapt) 100 procent! Nu is les 2 vrij gekomen."

(32) Comment / recommendation by participant

Participant	T5
Page	Lesoverzicht → 1. Je allereerste factuur → 7. Samenvatting

Description

Participant makes a comment about a sentence in the summary of the lesson ("that wasn't so difficult, yet, right?"). The participant notes that this sentence might be offensive to users that did have trouble completing the lesson. He advices us to be careful with these kinds of remarks.

Severity

○ Minor ⊙ *Serious* ○ Critical

(33) Answer to open-ended question by test facilitator

Participant T1

Page Lesoverzicht → 2. Tweede factuur (met btw) → 2. Opdracht

Description

Q: "Je kijkt nu een beetje verward. Alsof je het niet helemaal snapt, of ...?"

A: "Ja... omdat de vorige keer dat ik op enter drukte, verstuurde ie direct de factuur, en ik wil nu weer op enter drukken. Dus nu denk ik: zal ik het doen, zal ik het niet doen?"

Q: "Oke."

A: "Ik doe het wel."

(melding verschijnt boven factuur)

"Ah! Hey, nu krijg ik een melding dat ik de factuur nog niet helemaal juist heb ingevuld! Ha, dat is leuk."

Opmerking onderzoeker: Na korte discussie over Excel blijkt dat ze gewend is en het fijn vindt om met tab en enter te navigeren. Ze kan waarderen dat er ook toetsenbordondersteuning is in de factuuropdracht, (bijvoorbeeld het kunnen kiezen van een optie in de dropdown met de pijltjestoetsen). Maar de enter-toets werkt niet helemaal zoals verwacht. Ze wil niet dat er dan direct iets gebeurt (namelijk: het versturen van de factuur).

Severity

⊙ *Minor* ○ Serious ○ Critical

(34) Comment / recommendation by participant

Participant T1

Page Lesoverzicht → 2. Tweede factuur (met btw) → 2. Opdracht

Description

"(...) En dan kan ik in alle rust op 'verzenden' drukken. Hmm. Ik zou verwachten dat die hier staat. (wijst naar rechtsonder de factuur). Want je begint hier (wijst naar hoek linksboven van factuur) en eindigt daar (rechtsonder).

Severity

⊙ *Minor* ○ Serious ○ Critical

(35) Observation about pathways participants took

Participant T3

Page Lesoverzicht → 2. Tweede factuur (met btw) → 2. Opdracht

Description

Participant intuitively scrolls down after dragging the attachment to the dropzone, with no further comment or issues.

(36) Comment / recommendation by participant

Participant T4

Page Lesoverzicht → 2. Tweede factuur (met btw) → 2. Opdracht

Description

Participant indicates that it might make more sense for the buttons to be positioned below the invoice.

Severity

⊙ *Minor* ○ Serious ○ Critical

(37) Comment / recommendation by participant

Participant T4

Page Lesoverzicht → 2. Tweede factuur (met btw) → 2. Opdracht

Description

Participant indicates that he thinks a lot of concepts are introduced at the same time in the explanation that is shown.

Severity

○ Minor ⊙ *Serious* ○ Critical

(38) Comment / recommendation by participant

Participant T2

Page Lesoverzicht → 2. Tweede factuur (met btw) → 3. Opdracht

Description

Gebruiker laat blijken dat het niet meteen duidelijk is wat het verschil is tussen de balans van de toelichting sectie 2 en sectie 3 (debiteuren versus bankrekening. "Misschien is het voor de duidelijkheid beter om hier dan ook "Debiteuren: 0" neer te zetten".

Severity

⊙ *Minor* ○ Serious ○ Critical

(39) Observation about pathways participants took

Participant T4

Page Lesoverzicht → 2. Tweede factuur (met btw) → 3. Opdracht

Description

Participant wonders what the balance sheet on the previous page looks like in comparison to the one shown on the current page. Participant goes back to previous page to compare. After a little while the participant figures out the difference ("Bank account" vs "Debitors").

Suggestion by participant: perhaps you could highlight the difference.

Severity

⊙ *Minor* ○ Serious ○ Critical

(40) Comment / recommendation by participant

Participant	T2
Page	Lesoverzicht → 2. Tweede factuur (met btw) → 4. Opdracht

Description

"Het is voor mij heel erg duidelijk."

"Helemaal duidelijk. Goede uitleg."

(vervolgens alle drie meerkeuzevragen goed beantwoord.)

Opmerking onderzoeker: Later blijkt dat ze alles wat in les 1 en 2 aan bod is gekomen al eerder in workshops heeft geleerd (weliswaar een heel tijdje terug). Ze zegt dat het daarnaast lijkt dat boekhouden simpel is terwijl er wel een aantal randgevallen zijn die hier niet aanbod komen. (Nog niet). Daarnaast: "Maar ook als je net begint met boekhouden dan is dit gewoon hardstikke duidelijk. Je moet het gewoon goed lezen."

(41) Comment / recommendation by participant

Participant	T2
Page	Lesoverzicht → 2. Tweede factuur (met btw) → 7. Meerkeuzevraag

Description

"Ik ben wel benieuwd wat er gebeurt als ik een verkeerd antwoord kies. Of die dan weer terug gaat naar de (voorgaande) opdracht, of dat die dan een uitleg geeft waarom het antwoord fout is. Dan verwacht ik niet het 'ping' geluidje maar een bluh-geluidje. En dan 'fout', en dan (... uitleg).".

(42) Answer to open-ended question by test facilitator

Participant T2

Page Lesoverzicht → 2. Tweede factuur (met btw) → 7. Meerkeuzevraag

Description

Q: "Wat zou je verwachten dat er gebeurt nadat je het foute antwoord hebt gekozen?"

A: "(...) Maar niet dat je mensen nog weer het goede antwoord laat aankruisen want dat zou een beetje kinderachtig zijn. Dat is misschien leuk voor kinderen, maar (...)"

Opmerking onderzoeker: opvallend omdat later twee deelnemers juist zeggen dat dit wel een goed idee is, dat je er op die manier echt van leert.

(43) Comment / recommendation by participant

Participant T6

Page Lesoverzicht → 2. Tweede factuur (met btw) → 7. Meerkeuzevraag

Description

Participant notes the (educational) value of the multiple choice question explanations and the value of having to select a second answer when the first one was incorrect.

(44) Comment / recommendation by participant

Participant T6

Page Lesoverzicht → 3. Inkomende facturen → 1. Introductie

Description

Participant wonders whether it is possible to deduce what the number of the current lesson is (lesson #3), aside from looking at the URL and aside from going back to the previous screen.

Severity

⊙ *Minor* ○ Serious ○ Critical

(45) Observation about pathways participants took

Participant T4

Page Lesoverzicht → 3. Inkomende facturen → 2. Opdracht

Description

Participant attempts twice to single click on the attachment and notices that nothing happens.

Participant clarifies that he might feel it to have made more sense if the attachment opens when double clicking it. Participant is unsure of the added value of dragging and dropping the attachment.

Severity

⊙ *Minor* ○ Serious ○ Critical

(46) Problem encountered

Participant T5

Page Lesoverzicht → 3. Inkomende facturen → 2. Opdracht

Description

It is unclear to the participant that the attachment has been "uploaded". The participant notes that it would have been helpful if an icon or checkmark was displayed to show that the attachment is in fact "placed into" / "uploaded" into the target area.

Severity

⊙ *Minor* ○ Serious ○ Critical

(47) Observation about pathways participants took

Participant T6

Page Lesoverzicht → 3. Inkomende facturen → 2. Opdracht

Description

Participant drags attachment into dropzone. And waits for the upload animation to finish. Participant waits and pauses for a fraction of a second. Nothing seems to change. The participant scrolls down and continues the exercise.

Test facilitator interrupts and asks about the slight pause. The participant clarifies that he had expected the focus of the screen to go to the document below. And he notices that he had expected at least some visual change to occur after dropping the attachment into the dropzone.

Severity

⊙ *Minor* ○ Serious ○ Critical

(48) Problem encountered

Participant T6

Page Lesoverzicht → 3. Inkomende facturen → 2. Opdracht

Description

Participant enters the wrong amount (363 euros instead of 300 euros) in the input field. He clarifies that he thinks it would have made more sense if the amount was 363 euros of which 21 percent is value-added tax, instead of the other way around.

Severity

⊙ *Minor* ○ Serious ○ Critical

(49) Comment / recommendation by participant

Participant T6

Page Lesoverzicht → 3. Inkomende facturen → 3. Opdracht

Description

Participant suggests that it may come in handy to also show the before version of the balance sheet. Perhaps by showing it in an collapsed state where the user can click on it to expand the balance sheet and compare it to version already shown in the explanation. Participant notes the balance sheet should be collapsed initially to prevent too much information being shown at once / initially.

(50) Comment / recommendation by participant

Participant T4

Page Lesoverzicht → 3. Inkomende facturen → 6. Meerkeuzevraag

Description

Participants wonders why no multiple choice question answer is available for "decreased by 21 euros" (as opposed to "increased by 21 euros").

Severity

⊙ *Minor* ○ Serious ○ Critical

(51) Comment / recommendation by participant

Participant T6

Page Lesoverzicht → 3. Inkomende facturen → 6. Meerkeuzevraag / 7: Samenvatting

Description

Participant notes that, after answering these three multiple choice questions, he really feels like he understands the concepts explained in this lesson and that he can see the full picture now. He repeats how he really feels like he has learned something by completing these lessons.

D. Multiple choice question results

Performance by test group users on the various multiple choice questions available in the learning environment.

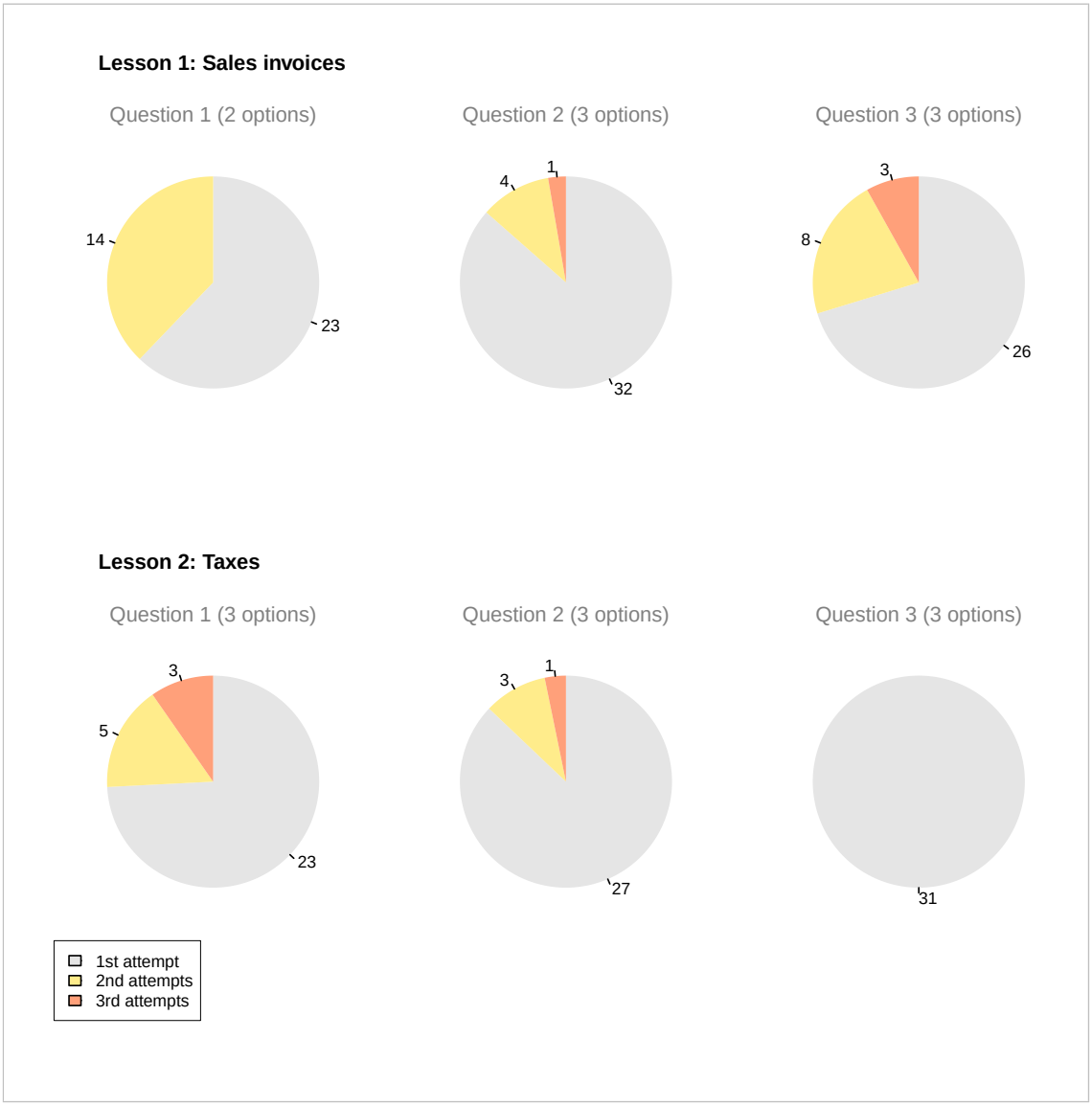


Figure D.1.: The distribution of number of attempts needed by test group users to answer multiple choice questions of lesson 1 and 2.

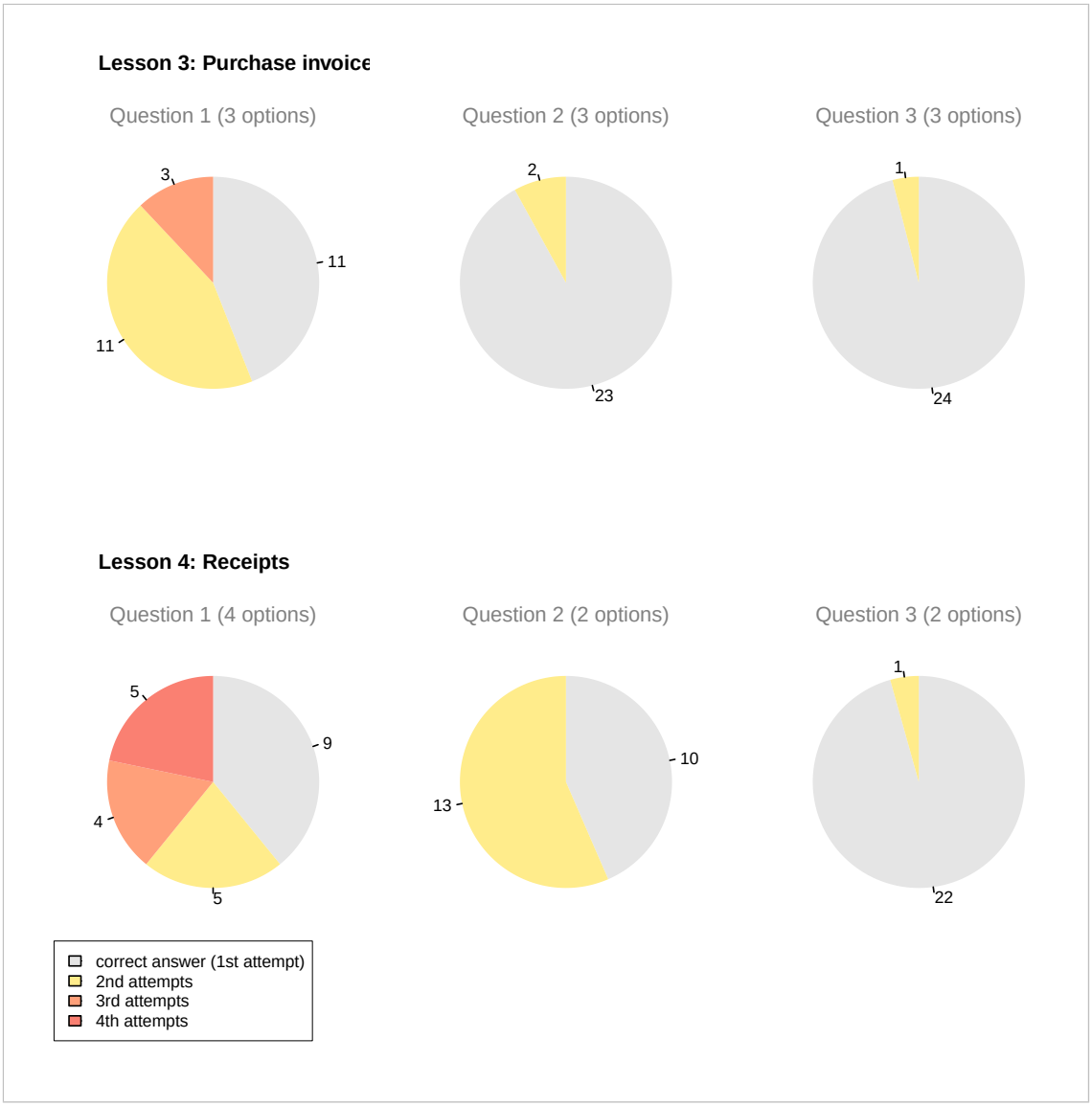


Figure D.2.: The distribution of number of attempts needed by test group users to answer multiple choice questions of lesson 3 and 4.